Bibliographical reference

Effects of Five Speech Masking Sounds on Performance and Acoustic Satisfaction. Implications for Open-Plan Offices, A. Haapakangas, E. Kankkunen, V. Hongisto, P. Virjonen, D. Oliva and E. Keskinen, Acta Acustica **vol. 97** (Number 4), 2011, pp. 641-655

DOI
https://doi.org/10.3813/AAA.918444

Acta Acustica united with Acustica
Vol. 97 (2011) 641 – 655

DOI 10.3813/AAA.918444

Effects of Five Speech Masking Sounds on Performance and Acoustic Satisfaction. Implications for Open-Plan Offices

A. Haapakangas[1)], E. Kankkunen[2)], V. Hongisto[1)], P. Virjonen[1)], D. Oliva[1)], E. Keskinen[2)]

[1)] Finnish Institute of Occupational Health, Indoor Environment Laboratory, Acoustics, 20520 Turku, Finland. valtteri.hongisto@ttl.fi

[2)] University of Turku, Department of Psychology, 20014 Turku, Finland

Summary

The aim of this study was to compare different sounds which can be used in open-plan offices to mask distracting speech. Fifty-four subjects were tested in seven sound conditions: speech, silence and five masked speech conditions. The five masking sounds were filtered pink noise, ventilation noise, instrumental music, vocal music and the sound of spring water. They were superimposed on speech. The masked speech conditions corresponded to an acoustically excellent open-plan office in respect to the Speech Transmission Index (STI 0.38). The speech condition (STI 0.62) corresponded to the STI obtained between nearby workstations in an acoustically poor open-plan office. Silent condition (STI 0.00) corresponded to the STI measured between two nearby private office rooms. In each of the seven sound conditions, the subjects performed a short-term memory task, a proofreading task and a creative thinking task and completed a questionnaire on acoustic comfort. Compared to silence condition, short-term memory performance deteriorated in speech condition and in most masked speech conditions. Compared to speech condition, performance improved when speech was masked with spring water sound. Ratings of acoustic satisfaction and subjective workload showed that masked speech conditions subjectively improved the working conditions compared to speech condition. Overall, the performance results and subjective perceptions showed that the spring water sound was the most optimal speech masker whereas vocal music produced negative effects similar to those of speech. The use of constant masking sounds should be preferred in open-plan offices instead of instrumental or vocal music.

PACS no. 43.55.Hy, 43.66.Dc, 43.66.Lj

1. Introduction

1.1. Noise in open-plan offices

The concept of open-plan offices and landscaped offices was introduced in the 1960's. Nowadays, most office workstations are located in open-plan offices. However, a growing body of scientific evidence points at the disadvantages of open-plan offices, such as worker dissatisfaction and increased cognitive workload (e.g. [1]), fatigue and difficulties in concentration [2, 3], and subjective impairment of work performance (e.g. [3, 4]. The most commonly mentioned cause for these problems is the acoustic environment of open-plan offices, i.e. office noise and lack of speech privacy [5, 6, 7]. In office surveys, coworkers' speech is typically rated as the most distractive noise source [3, 4, 7].

An appropriate room acoustic design for open-plan offices requires that the distraction by speech sounds is sufficiently controlled. In order to achieve this, room acoustic guidelines encourage the simultaneous consideration of several factors: absorbers on ceiling, walls and furniture, high screens and storage units, distance between workstations, enclosure of workstations and the use of artificial masking sound [8, 9, 10]. All these factors, except the last one, aim at reducing the sound pressure level (later: level) of speech. Masking sound is needed to cover the remaining speech sounds to reach negative values of speech-to-noise ratio. The speech-to-noise ratio, L_{SN}, is defined as the difference of the levels of speech and background noise.

The necessity of masking sound was already recognized at the early stage of the development of the open-plan office concept [11, 12]. Artificial masking sound is recommended if the inherent A-weighted background noise level, $L_{\mathrm{A,eq}}$, in the office is below 40 dB. This is very often the case as the noise levels of air conditioning and office equipment have been reduced. This study focuses on the question, what kind of audio material should, or should not, be used as masking a sound.

1.2. Effects of irrelevant speech on cognitive performance

The need for the acoustic control of speech sounds is backed up by ample evidence from cognitive psychology.

Received 19 August 2009,
accepted 2 April 2011.

Background speech has an effect on several cognitive abilities, with the performance decrements varying between 4 and 41% in different laboratory experiments, see review by Hongisto [13]. The best-established finding in this area of research is the disruption of serial memory by irrelevant sound (e.g. [14, 15], known as the irrelevant sound effect, ISE. In a typical experiment, subjects are required to recall visually presented verbal material, e.g., digits, in the order of presentation, while ignoring the background sounds. This is a classic procedure for testing short-term memory, which is a type of memory used for maintaining information in mind for a short period of time so that further processing of the information is possible. Short-term memory is a central component in human information processing. The disruption of short-term memory by irrelevant background speech is empirically robust. Most individuals show the effect and the average impairment of performance can be as high as 30% [16]. Irrelevant background speech has also been found to disturb more complex tasks, such as proofreading [17, 18], reading comprehension [19], activation of prior knowledge in long-term memory [20] and logical reasoning [21].

There are several explanations for the disruption of performance caused by irrelevant speech. Currently, the most common view seems to be that performance is disrupted if the cognitive processes required in the task conflict with those processing auditory stimuli (*interference-by-process view*, e.g. [16, 22]). In other words, there is no general effect of noise on performance but it rather depends on both task demands and the properties of the sound. The ISE observed in serial recall is generally attributed to the acoustic variability of the sound (*changing-state hypothesis*, see [23]), and it does not occur with continuous noise, like pink noise, where temporal variations cannot be distinguished [24]. The disruption is only expected to occur if the sequences of sounds sufficiently change in frequency, pitch or modulation. The automatic processing of such changing sounds is assumed to engage the same cognitive processes that are required in maintaining order information in short-term memory [22]. The level of background speech does not affect performance within the range from 35 to 85 dBA provided that speech intelligibility is perfect [25, 26, 27].

In addition to speech, the ISE occurs with other variable sounds, such as irrelevant tones that are presented with the same duration and frequency as syllables in speech [28] and music [29], particularly vocal music [30, 24]. The changing-state features seem to be the deciding factor, i.e. instrumental music has to have distinct temporal-spectral variations to cause a disruption [31].

The changing state hypothesis presumes that when speech is the irrelevant sound, the information content of speech does not contribute to the effect observed in serial recall performance. However, there is some contrary evidence suggesting that the semantic processing of speech brings additional disruption. LeCompte *et al.* [32] showed that meaningful speech disrupts serial recall more than irrelevant tones used by Jones and Macken [28]. Words with negative affective content also produce a greater ISE than neutral words [33]. Following from the interference-by-process view, the semantic content of irrelevant speech, e.g., whether it is native or foreign language, is expected to be most critical in tasks that require semantic processing, such as reading comprehension or proofreading.

1.3. Speech Transmission Index

It has been suggested that the effect of speech on cognitive performance increases with increasing speech intelligibility [13, 20, 25, 26, 34, 35]. Acoustic satisfaction also decreases with increasing speech intelligibility [18, 20, 35, 36]. Compared to the changing state theory, speech intelligibility is a more suitable concept for explaining speech-related performance effects in open-plan offices where speech sounds are the dominant source of distraction and speech intelligibility can be objectively estimated.

The Speech Transmission Index, STI, is widely used in Europe as an objective descriptor of speech intelligibility (IEC 60268-16). It can easily be measured between speaker and listener workstations in open-plan offices. The measurement of STI between workstations in also included in the international committee draft standard which describes the measurement of room acoustics in open-plan offices (ISO/DIS 3382-3) The STI is a rough estimate of the percentage of correctly heard consonant-vowel-consonant syllables. When STI is 0.00, subjective speech intelligibility is zero. It is reached when the room reverberation is large and/or speech-to-noise ratio, L_{SN}, is below −15 dB. When STI is 1.00, subjective intelligibility is perfect. It is reached only in anechoic conditions when the L_{SN} is larger than +15 dB. In open-plan offices, the STI mostly depends on the L_{SN}. Therefore, speech privacy can be improved by reducing the L_{SN}. Typical values of STI in open-plan offices are between 0.00–0.80 depending on the room acoustic conditions and the distance to the speaker [10].

Hongisto [13] has presented a model that describes a relation between the STI and cognitive performance. The model was based on a literature review of laboratory experiments that have investigated the effect of speech on performance. The model predicts that performance starts to decrease when the STI exceeds 0.20. The maximum decrease is reached when the STI exceeds 0.60. The model gives clear evidence that the STI should be minimized between workstations to maintain optimum cognitive performance.

1.4. Quality of masking sound

The use of masking sound decreases speech intelligibility and should similarly decrease the performance effects of background speech. But what should the level and the spectrum of masking sound be like? Previous studies have typically used pink noise (e.g. [34]), pink noise filtered to approximately −5 dB/octave [18, 20] or white noise [37] as a masking sound. The spectra are presented in the

Methods. Veitch *et al.* [36] conducted two laboratory experiments on the effects of the spectrum and the level of continuous and random electrical noise on acoustic satisfaction and speech masking efficiency. They concluded that an optimum spectrum of masking sound should approximately follow the speech spectrum, i.e. a slope of −5 dB per octave increment in the frequency range 125–8000 Hz, and the level should not exceed 45 dBA. This kind of sound is easy to produce by filtering pink noise.

Music could also be one potential masking sound instead of continuous sounds. Music has been present in workplaces since the 20th century [38]. It is spontaneously used by many workers to mask unwanted sounds of the office environment [3]. Music is believed to have a mood-lifting influence on workers [39, 40], which continuous noise has not been assumed to do. However, the more the work requires high concentration, the more disturbing background music is experienced by the workers [39].

Schlittmeier and Hellbrück [41] have tested the potential use of music as a masking sound. The laboratory experiment included two types of instrumental music: staccato music, which had distinct temporal-spectral variations, and legato music that lacked strong changing-state features. Music conditions were compared to pink noise (see Methods for spectrum). All three masking sounds, presented at $L_{\mathrm{A,eq}} = 55\,\mathrm{dB}$, were superimposed onto recorded office noise, also presented at $L_{\mathrm{A,eq}} = 55\,\mathrm{dB}$. Only pink noise improved the performance on a serial recall task when compared to the office noise without masking. Subjective ratings revealed that most participants would prefer legato music to pink noise in open-plan offices. We suggest that the pink noise might have been assessed more favourably if a more comfortable spectrum had been used, e.g., pink noise filtered to −5 dB per octave suggested by Veitch *et al.* [36]. Pink noise possesses equal level at every octave band, while comfortable masking follows the spectrum of speech, including significantly less high frequency sounds (see Methods for spectrum).

The present literature does not directly answer the question what kind of masking sound would be the most appropriate for open-plan offices with respect to both work performance and acoustic satisfaction. The most common inherent masking sounds in offices are ventilation noise or external constant road traffic noise. However, they are inappropriate in most cases because of improper spatial uniformity, sound pressure level or sound quality. Non-speech sounds caused by working and babble can also produce appropriate sound masking in large and reverberant offices. Babble can be formed by the simultaneous talk of more than eight people which makes individual speech undistinguishable. Jones and Macken [42] showed that babble does not cause as a strong distraction that is caused by one speaker or a couple of simultaneous speakers. However, the level of babble cannot be controlled, and when too loud, it will produce problems for normal conversation. Therefore, the promotion of babble does not belong to the room acoustic design guidelines. In conclusion, artificial masking sounds, such as optimally filtered pink noise, or recorded sounds from the nature or the environment, have been of particular interest among product manufacturers and end-users. Artificial masking sound is produced by loudspeakers. Commercial masking systems exist both for local and department level purposes. The former is suitable for situations where each workstation needs to be provided with an independent masking sound system. The latter is suitable for situations where the whole open-plan office needs to be exposed to the same masking sound.

The application of artificial masking sound technology is often regarded with reservations among architects, workplace representatives and workers, if they are not familiar with the principles of speech masking. Therefore, experimental studies should investigate both positive and negative effects of masking sound to reach a better impact on workplace design. One of the greatest concerns is related to the adverse health effects of noise as such. Increased occupational and environmental noise is associated with interference of sleep, cardiovascular and psychophysiological systems and social behaviour, and with annoyance responses [43]. However, masking sound increases the A-weighted level, $L_{\mathrm{A,eq,8h}}$, in the office only marginally. The level of office noise is usually between 46 dB and 58 dB [44]. The recommended level of masking sound is usually 45 dB at most [8]. Adding such a masking sound over office noise increases the overall level less than 1 dB. Such a difference should not increase the physical load.

1.5. Aim of the study

The aim of this study was to compare five different masking sounds in terms of cognitive performance and subjective perception. Similar sounds can be used to mask distracting speech in many open-plan offices. The study included seven sound conditions: silence, speech and five masked speech conditions. The different types of masking sounds and the correspondence of the sound conditions and open-plan office conditions are described in Table I. The performance effects were assessed with three tasks: serial recall, proofreading and creative thinking tasks, which represent different aspects of every-day office work. The following research questions were addressed.

Q1. Do speech (STI 0.62) and masked speech (STI 0.38) impair cognitive performance and acoustic satisfaction in comparison to silence (STI 0.00)? Silence is used as a baseline condition as it is assumed to represent an ideal working environment for cognitively demanding work. It is hypothesized that speech will impair both performance and acoustic satisfaction. A similar effect is also expected to occur in masked speech conditions.

Q2. Does the reduction in STI improve cognitive performance and acoustic satisfaction, i.e. do the five masked speech conditions differ from speech? It is hypothesised that a lower STI will improve both task performance and acoustic satisfaction.

Q3. Do the five masked speech conditions differ from each other in terms of their benefits or adverse effects regarding cognitive performance and acoustic satisfaction?

Table I. Acoustic descriptions of the seven sound conditions. All levels represent equivalent A-weighted sound pressure levels. L_S = speech, L_N = masking, L_{tot} = total level, L_{SN} = speech-to-noise ratio, i.e. the difference of speech and masking. STI = Speech Transmission Index. The following abbreviations are used in figures:

"silence": speech and masking sounds absent,
"speech": highly intelligible speech, no masking sound,
"pink": speech masked with filtered pink noise,
"ventilation": speech masked with ventilation noise,
"instrumental": speech masked with instrumental music,
"vocal": speech masked with music containing lyrics,
"water": speech masked with spring water sound.

abbreviation	L_S	L_N	L_{tot}	L_{SN}	STI
silence	–	37.5	37.5	$-\infty$	0.00
speech	47.8	37.5	48.2	10.4	0.62
pink	44.8	46.1	48.5	−1.3	0.39
ventilation	44.8	45.0	47.9	−0.2	0.37
instrumental	44.8	45.1	47.9	−0.3	0.35
vocal	44.8	45.1	47.9	−0.3	0.36
water	44.8	45.0	47.9	−0.2	0.40

If such differences were found, it would suggest that the masking sound type is important, in addition to the acquired STI value. Also, for a masking sound to be optimal, it should not be rated as annoying in itself.

2. Methods

2.1. Subjects

Fifty-four university students, 33 men and 21 women took part in the study. Subjects were 19 to 45 years old (mean = 26; standard deviation = 5) and native Finnish speakers. None of the subjects had dyslexia or reported any hearing difficulties. Subjects were recruited via the university's email lists and were paid 50 euros for participating.

2.2. Laboratory room

The research was carried out in an office laboratory (30 m^2) of the Finnish Institute of Occupational Health in Turku (Figure 1). The room was designed to resemble a neutral clerical office. The most important factors of the indoor environment were designed to meet current recommendations. The workstations were separated on three sides by 1.30 meter high screens. The furniture and computers complied with present-day ergonomic recommendations.

An indirect lighting system was used. Illuminance was set at 400 lx on the table of occupied workstations. No glare problems occurred in the workstations. The equivalent octave band level, temperature, CO_2 concentration of the room, supply airflow rate to the room and relative humidity were monitored (Sinus Harmonie Light). Room temperature was measured at the height of hands and at the height of feet. The room temperature rose by 1 to 2 °C during the experiments, but remained within comfortable range from 22 to 25 °C. The supply airflow rate to the room was 60 l/s. The relative humidity varied between individual tests from 18% to 26% depending on the weather. The CO2 concentration in the room (520–730 ppm) did not produce a risk of fatigue. The A-weighted equivalent level was monitored in each session to check that the level did not exceed the nominal values of Table I. Neither the subjects nor the researcher affected the average levels of the room by their own activity. The average levels in the experimental condition deviated from the nominal levels presented in Table I by less than 1 dB.

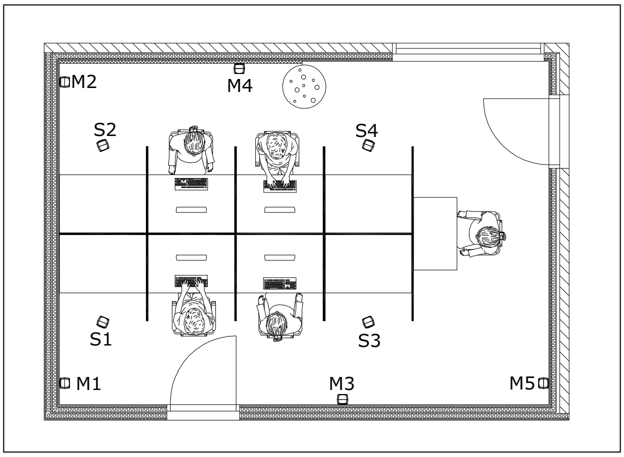

Figure 1. The office laboratory contained eight workstations. Four subjects sat in the middle. Speech samples were produced randomly from four dummy workstations (S1-S4). Masking sound was produced by five loudspeakers hidden above the suspended ceiling (M1-M5).

The early decay time and speech-to-noise ratios were determined (winMLS) to define the STI of normal effort speech between the speech loudspeakers and occupied workstations [10]. The average early decay times were 0.46, 0.40, 0.35, 0.35, 0.30, 0.28 and 0.26 in the octave bands 125, 250, 500, 1000, 200, 4000 and 8000 Hz, respectively. The values correspond to an open-plan office with very high room absorption.

2.3. Design of sound conditions

The main motive of this study was to compare different masking sounds. In commercial applications, filtered pink noise has been most frequently used as the speech masking sound, because it is easily available, inexpensive, stable, maintenance-free and an effective speech masker. It was also used in our previous laboratory and field studies [18, 20, 45, 46]. Several workplace designers and end-users have suggested the use of natural sounds, music and ventilation as they might be more acceptable than artificial sounds, like filtered pink noise. However, most natural sounds do not mask speech very well because of spectral features. We decided to use recorded spring water sound because its spectrum was close to the preferred masking sound spectrum suggested by Veitch *et al.* [36]. Both vocal music and instrumental music were selected because radio listening by earphones is a very popular way to personally mask surrounding speech. Finally, we selected recorded

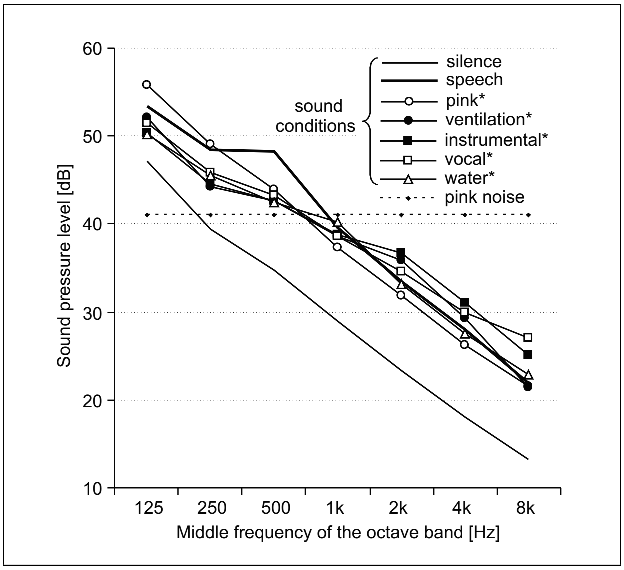

Figure 2. Octave band spectra of the seven sound conditions recorded in the room. The curves represent the time average in the workstations. Spectra represent total sound pressure levels of Table I. The pink noise spectrum is shown to demonstrate the difference to the filtered pink noise. The conditions with an asterisk (*) are masked speech conditions in which the sound was presented together with speech (see Table I).

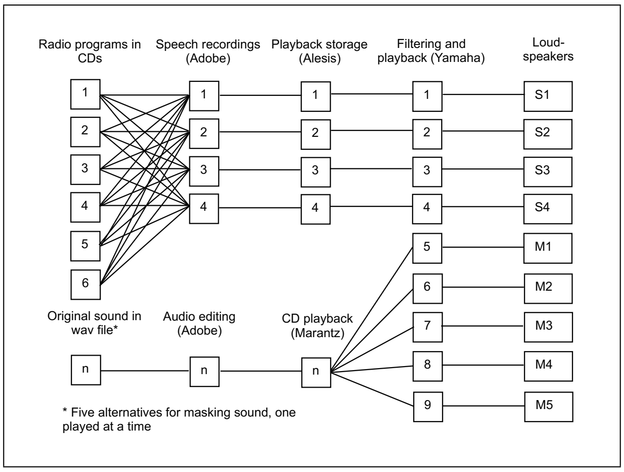

Figure 3. Creation of speech (S) and masking (M) sounds.

ventilation sound as it is the most common environmental sound in buildings. No other masking sounds were selected because of the limitations of the experimental design.

The spectra of the five masking sounds, as well as the spectra of bare speech and silence, are presented in Figure 2. Because the spectra were measured in the room, they include the background noise of the room. Therefore, the bass content of speech condition is higher than the speech actually emitted by the loudspeakers and the spectrum deviates slightly from typical standardized unisex speech spectrum [47]. In addition, the room modes and room acoustics may change the spectrum of speech slightly.

The sound conditions were designed to correspond to acoustic conditions in offices in terms of level and STI. This study contained three STI values, 0.00, 0.38 and 0.62. Silence condition corresponds to STI 0.00. Such an environment is reached in private office rooms when doors are closed and the walls are sound-proof. An STI of 0.38 represents an open-plan office with very good acoustic design where speech intelligibility is significantly lowered. An STI of 0.62 represents a poorly designed open-plan office where speech intelligibility is high. Both STI values can be reached between nearby workstations in open-plan offices depending on the room acoustic design [9, 10].

The level of normal effort speech is approximately $L_{\mathrm{A,eq,S}} = 59\,\mathrm{dB}$ at a distance of 1 meter from the mouth [47]. According to the laboratory experiments of Virjonen *et al.* [9], the A-weighted level of normal effort speech, measured in the neighbouring workstation behind the screen, varied between 39 and 55 dB, depending on room absorption and screen height. Hongisto *et al.* [48] and Virjonen *et al.* [10] have made similar measurements in open-plan offices. However, the level of speech never fell below 45 dB in the nearest workstation. Therefore, the speech levels of 45 dB (during the five masked speech conditions) and of 48 dB (during speech condition) were justified on the ground of this field evidence.

During the speech condition, no masking sound was played. The only speech masker was the background noise of the room, which corresponded with the silent condition. However, its level was 10 dB lower than the level of speech. Therefore, the speech intelligibility was high.

The reduction of the STI was done by reducing the speech-to-noise ratio, see Table I. The total A-weighted level was 48 dB in all sound conditions except the silent condition. A constant total level was used since level differences were beyond the scope of this study. The masked speech conditions consisted of two modifications to the speech condition: the level of speech was reduced by 3 dB to 45 dB. This reduction of speech level corresponds with a minor room acoustic improvement. In addition, the level of masking sound was presented at the level of 45 dB.

2.4. Creation of speech and masking recordings

Masking sound and speech were presented in the laboratory room from separate loudspeaker systems, see Figures 1 and 3. Six speech recordings were edited because all sound conditions except silence included speech sounds (Adobe Audition). Each of them was 25 minutes long. Office sounds other than speech were not used. Materials for the recordings were obtained in CD format from eight different radio programs (YLE, Finland). The audio materials were stored before the broadcasting so that the speech was uncompressed and contained normal level variations of speech. The programs were in interview format with one male and one female native Finnish speaker taking part. The programs were intentionally selected to be as neutral as possible, and were not of current interest nor directed at the age group of the subjects. The speech was cut into 10- to 30-second-long full-sentenced samples. In each sample, only one person was talking. The levels were normalized between samples. There was a 3.2-8.8-second silence between each sample. The successive samples were taken from different programs so that there was no plot to follow. Thus, the final recording resembled a normal office environment where both speech and silence alternate arbitrarily. The speech samples were mostly the same as in Haka *et al.* [20]. Final speech recordings were stored randomly on 4 channels of a hard disc recorder (Alesis adat HD24) and distributed on four separate loudspeaker channels. Only one loudspeaker produced speech at a time. The speech recordings were played via four loudspeakers (Genelec 8020) that were located in neighboring workstations around the subjects at a height of 120 cm. A statistical analysis of the variation in the level and STI was made on the basis of 30-second-long samples. The standard deviation of the A-weighted level of speech was ±1.4 dB during the whole speech recording (Table I), resulting in a standard deviation of the STI of about 0.05. When any of the four speech loudspeakers was active, the variation in level was within ±1 dB in the four subjects' workstations. Low standard deviation was obtained because the ceiling above the workstations was reflecting so that the speech was easily distributed over the workstations.

The vocal music consisted of five Finnish pop-songs frequently heard on the radio ('Suomen neito' performed by Jani Wickholm; 'Virta mua kuljettaa' by Lauri Tähkä and Elonkerjuu; 'Mansikkamäki' by Katri Ylander; 'Enkelten kaupunki' by Jippu; 'Ruoste' by CMX; 'Myrskypääskynen' by J. Karjalainen; 'Unohtaisinpa' by Kristiina Brask). Instrumental music consisted of five classic pop and rock pieces interpreted by the London Symphony Orchestra ('She's not there' by Rod Argent; 'Don't give up' by Peter Gabriel; 'You're the voice' by Reid, Thompson, Ryder and Qunta; 'Lady in red' by Chris de Burgh; 'Separate lives' by Stephen Bishop).

Masking sounds were obtained either from live recordings (ventilation and spring water), from commercial audio tracks (for instrumental and vocal music) or from digital sources (filtered pink noise). The sounds were transferred to a mixing program (Adobe Audition 1.0) and stored on CD's. The CD was played in the experiments (Marantz CD4000) and filtered by a 16-channel digital mixer (Yamaha DME24N) before feeding to the loudspeakers. The filtering was done in 1/3-octave bands to compensate the slight spectrum changes caused by the room. Masking sounds were played via 5 loudspeakers (Genelec 8020) mounted above a suspended ceiling. Each of the five masking loudspeakers had an independent filter (Yamaha DME24N) to obtain an equal frequency response in each workstation.

2.5. Questionnaire

At the beginning of the experiment, background information including gender, age, the amount of sleep during the preceding night, and arousal, were measured. Arousal was measured with a modified version [49] of the Karolinska sleepiness scale [50]. An acoustic satisfaction questionnaire was administered after each sound condition. This included 16 items that measured different subjective aspects of the sound condition (Table II). Most of the questions were taken from the study by Haka *et al.* [20]. The items also included four future-oriented questions (items 9–12 in Table II) adapted from Veitch *et al.* [36] that evaluated the applicability of the sound environment in everyday working. Attempts were made to group items together with factor analysis but the factor structures differed between the sound conditions, i.e. sum variables could not be formed using statistical methods. Therefore, two sum variables were created on the basis of theoretical consideration (Table II). The sum variable of greatest interest was *acoustic satisfaction* which included 12 items and had Cronbach's Alphas between 0.91 and 0.94 in different sound conditions. Four items focusing on subjective ratings of performance and effort were averaged together to form a sum variable for *subjective workload* (Cronbach's Alphas between 0.72 and 0.85). This variable was included because previous studies have suggested that the subjective awareness of distraction may encourage subjects to invest more effort in performance, which may obscure the relationship between sound condition and measured performance [26]. The questionnaire also included an assessment of arousal and ratings of disturbance caused by different noise sources. All the questionnaires were presented with an internet-based software (Digium, Finland). At the end of the experiment, the disturbance caused by other environmental factors was rated on a scale from 1 to 5 to control that they had not affected the experiment.

2.6. Performance tests

In the *serial recall task*, subjects had to recall visually presented digits from 1 to 9 in the order of presentation. The digits were presented in a random order, each digit appearing only once in a sequence. Digits were presented on the computer screen at the rate of 1 per second, with an interdigit interval of 1.5 seconds, i.e. the onset-to-onset of digits was 2.5 seconds. After a sequence, subjects recalled the digits by clicking numbers in the same order on a 3x3 array on the screen. Subjects had the possibility to click 'empty' if they could not remember a number in a certain serial position. The task contained 10 sequences. The percentage of digits not recalled in their correct positions was calculated for each sequence. The average score from all 10 sequences (mean error rate) was analysed.

The *creative thinking task* was a formulation [51] of Guildford's Alternate Uses Task [52]. The subjects were instructed to write down as many alternative uses for an object as possible in 5 minutes. They were presented with one name of an object on a paper and the common use of the object was given. The 7 objects were brick, car tire, broomstick, fishing net, fork, potato and tennis ball. Two dependent variables were formed for the analysis: ideational fluency and ideational originality. Ideational fluency was calculated by adding up the number of responses

Table II. Means and standard deviations (in brackets) for the sum variables of acoustic satisfaction and subjective workload, and for the individual items of the sum variables in the seven sound conditions. The conditions with an asterisk (*) are masked speech conditions in which the sound was presented together with speech (See Table I). SE = Sound environment, r = reverse-scored in the sum variable of acoustic satisfaction. Individual items were rated on a scale 1–5 (completely disagree – completely agree). ventil.: ventilation, instr.: instrumental, p.o.t: periods of time, consid.: considering.

Variable	silence	speech	pink*	ventil.*	instr.*	vocal*	water*
Acoustic satisfaction	4.5 (0.6)	2.3 (0.9)	3.0 (0.9)	2.9 (0.8)	2.8 (0.9)	2.3 (0.9)	3.0 (0.8)
The twelve items in acoustic satisfaction:							
1. SE was pleasant	4.4 (1.0)	2.2 (1.1)	2.7 (1.0)	2.3 (1.0)	2.7 (1.2)	3.0 (1.2)	2.8 (1.0)
2. SE was disturbing [r]	1.4 (0.9)	3.9 (1.2)	3.0 (1.2)	3.2 (1.2)	3.3 (1.3)	3.8 (1.1)	3.2 (1.1)
3. SE was acceptable	4.7 (0.6)	2.6 (1.2)	3.4 (1.1)	3.2 (1.0)	3.1 (1.2)	3.1 (1.2)	3.4 (1.0)
4. SE was loud [r]	1.1 (0.3)	3.4 (1.3)	2.7 (1.1)	2.8 (1.0)	2.9 (1.2)	3.4 (1.2)	2.7 (1.0)
5. Overall I was satisfied with SE	4.4 (1.0)	2.1 (1.1)	2.7 (1.0)	2.4 (1.0)	2.7 (1.2)	2.6 (1.1)	2.7 (1.1)
6. Habituation to SE was easy	4.7 (0.7)	2.6 (1.2)	3.6 (1.1)	3.2 (1.0)	3.2 (1.2)	3.1 (1.4)	3.5 (1.1)
7. Surprising changes occurred in SE [r]	1.1 (0.4)	3.0 (1.3)	2.6 (1.3)	2.5 (1.2)	2.9 (1.4)	2.8 (1.4)	2.3 (1.3)
8. SE caught my attention often [r]	1.7 (0.9)	3.9 (1.0)	2.9 (1.1)	2.9 (1.1)	3.4 (1.1)	4.0 (0.8)	2.9 (1.0)
9. I could work uninterrupted for long p.o.t.	4.2 (0.9)	2.2 (1.2)	2.9 (1.1)	2.7 (1.2)	2.4 (1.3)	2.3 (1.3)	2.9 (1.2)
10. I could work effectively for long p.o.t.	4.2 (1.0)	2.1 (1.1)	2.9 (1.1)	2.7 (1.1)	2.5 (1.2)	2.4 (1.1)	2.8 (1.2)
11. SE would not annoy me	4.2 (1.0)	2.2 (1.2)	2.6 (1.2)	2.5 (1.2)	2.5 (1.3)	2.6 (1.3)	2.6 (1.1)
12. I could accept SE consid. the work I do	4.4 (0.9)	2.3 (1.3)	2.8 (1.3)	2.7 (1.2)	2.5 (1.3)	2.7 (1.2)	2.8 (1.3)
Subjective workload	2.0 (0.8)	3.5 (0.7)	3.0 (0.8)	3.1 (0.8)	3.2 (0.9)	3.5 (0.8)	3.1 (0.9)
The four items in subjective workload:							
1. SE impeded my ability to concentrate	1.5 (0.9)	4.0 (0.9)	3.2 (1.0)	3.4 (1.2)	3.6 (1.2)	4.0 (0.9)	3.4 (1.1)
2. SE impaired my performance	1.4 (0.9)	3.7 (0.9)	3.0 (1.1)	3.2 (1.2)	3.4 (1.1)	3.7 (1.1)	3.2 (1.2)
3. The tasks felt difficult.	2.5 (1.2)	3.1 (0.9)	2.8 (1.0)	2.9 (1.0)	2.9 (0.9)	3.1 (1.1)	2.9 (1.1)
4. I had to exert myself to get the tasks done.	2.4 (1.2)	3.1 (1.2)	2.9 (1.2)	3.0 (1.2)	3.1 (1.2)	3.2 (1.2)	2.9 (1.2)

(uses). Ideational originality was measured by determining the C-score of each use. It is calculated by $C = \log_2(1/p)$, where p is the probability of the use among all subjects in a specified sound condition. For example, if only one subject of 54 subjects proposes a specific use, the probability is 0.0185 and the C-score becomes 5.75. High C-scores indicate greater ideational originality. The C-scores were added up for each subject in each sound condition. The normalized C-score of a sound condition was obtained by dividing the sum of all subjects' C-scores by the average C-score over all sound conditions.

The *proofreading task* was modified from the one used by Venetjoki *et al.* [18]. Subjects were advised to mark as many mistakes in a text as possible in 10 minutes. Each text was four pages long and had approximately 1150 words. Altogether seven equally complex versions were used. The texts were drawn from Ahonen *et al.* [53]. Each text included 60 mistakes of which half were non-contextual errors and the other half contextual errors. Previous research suggests that there are differences in how irrelevant sound affects the detection of non-contextual and contextual errors [17, 54]. Non-contextual errors included different types of misspellings, e.g., character substitutions and omissions of single letters. All these misspellings produced non-words. Contextual errors included grammatically incorrect formulations of words and words that did not fit the context. All these words were legitimate Finnish words. Thus, their detection required semantic processing of the text, whereas the non-contextual errors could be detected by visual search. The main dependent variables were the total number of correctly found non-contextual and contextual errors. Reading speed, determined by the number of lines read, was also analysed.

2.7. Experimental design

The study was a repeated measures design, i.e. all subjects took part in all seven sound conditions, thus acting as their own control. Silence and speech conditions were included as control conditions. There were five masked speech conditions containing a mix of speech and a specially designed speech masking sound, see Table I. The presentation order of sound conditions was a possible source of error in this type of design, as the effect of some extraneous factors, such as initial anxiety or fatigue, will be associated with different stages of the test session. To control this error, the conditions would typically be counterbalanced across subjects, i.e. presented in all possible orders with an equal number of subjects in each. This was not possible in this experiment because of the seven sound conditions and only 14 test groups. Thus, the sound conditions were presented in a quasi-random fashion with the constraint that every sound condition was presented twice as the first and twice as the last one, and approximately twice in every serial position between the first and the last. Moreover, care was taken that every masked speech condition occurred twice after the silent condition and twice after the speech condition. The speech recordings were divided between speech and masked speech conditions so that every recording was presented two to three times in each sound condition.

In each sound condition, subjects performed the serial recall, the creative thinking and the proofreading task. The task performance and the subjective perceptions of the sound conditions were the dependent variables in this study. The seven matching versions of proofreading and creative thinking tasks were counterbalanced across sound conditions so that every version was presented twice in each sound condition. Half of the subjects performed first the serial recall task which was followed by the proofreading and the creative thinking task (Group 1), and the other half performed first the proofreading task which was followed by the creative thinking and the serial recall task (Group 2).

2.8. Procedure

The experiments were conducted in May 2008. Half of the subjects were tested from 8 a.m. to 12.30 p.m. and the other half from 1 p.m. to 5.30 p.m. Three to four subjects were tested at the same time in 14 groups. Subjects were separated by screens and were not allowed to communicate with each other. The experimenter was present in the room during the whole experiment (Figure 1). Before the experiment started, subjects were informed that the study was investigating performance in an office-like environment. Subjects filled in the questionnaire gathering background information. This was followed by a 30-minute practise session of the tasks in silence. Before the actual experiment started, subjects were instructed to ignore the sounds and concentrate on the tasks. The experimenter switched on the masking and speech sounds, and then said which task the subjects should start. After the tasks had been completed, the experimenter turned off the sounds. Each sound condition lasted approximately 25 minutes, and was followed by the questionnaire (5 min) before the next sound condition began. Ten-minute breaks were given after the 1st and the 4th sound condition. Subjects were informed in detail about the aim of the study at the end of the experiment.

2.9. Statistical methods

SPSS 16.0 (SPSS Inc., Chicago, Il, USA) was used for the statistical analysis. The normality of the data was tested with the Kolmogorov-Smirnov test. The initial analysis of the performance data revealed that some of the results were affected by the presentation order of the tasks, i.e. whether the task had been performed in the beginning of the test condition or later on during the session. Thus, the serial recall task and the creative thinking task were analysed using a mixed 7 × 2 ANOVA with the sound condition as a within-subject variable and the presentation order as a between-groups factor (Groups 1 and 2). The proofreading task was analysed using a 7 × 2 × 2 ANOVA with the seven sound conditions and the two error types as within subject factors, and the presentation order as a between groups factor. F-values were tested on homogeneity of variance with Mauchly's test of sphericity. A repeated measures ANOVA was used for the questionnaire items that were normally distributed and the Friedman's test for the variables that were not. Paired comparisons between sound conditions were performed using t-tests for the variables that were normally distributed and Wilcoxon signed-rank test for the variables that were not. An alpha level of 0.05 was used in all analyses. Benjamini-Hochberg procedure [55, 56] was used for alpha-error adjustments in t-tests. The estimate of effect size was measured in variance analysis with partial Eta squared and in the t-tests with Cohen's d. The suggestive values of d for small, medium and large effect sizes are 0.20, 0.50 and 0.80, respectively [57]. Four subjects' data could not be recorded in the serial recall task due to a technical problem. One subject was excluded from the analyses of the proofreading task because she had misunderstood the instructions.

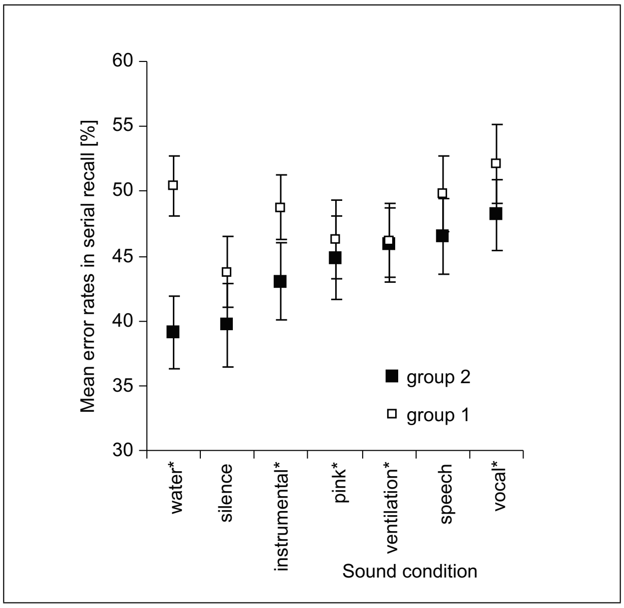

Figure 4. Mean error rates (%) and standard errors of means for performance in serial recall in the seven sound conditions. Group 1 always performed the task in the beginning of each sound condition whereas Group 2 performed the task as the last task in each sound condition. The research hypotheses were tested with Group 2. The conditions with an asterisk (*) are masked speech conditions in which the sound was presented together with speech (see Table I).

3. Results

The main effect of the sound condition on each variable is first reported. Research questions are then addressed by paired comparisons: i) silence is compared to speech and all masked speech conditions to address the first research question, ii) speech is compared to masked speech conditions to address the second research question, and iii) masked speech conditions are compared to each other to address the third research question.

3.1. Performance measures

The analysis of the serial recall task revealed a significant main effect of the sound condition for the mean error rates (Figure 4, $F_{6,288} = 5.09$, $p < .001$, partial $\eta^2 = .096$). However, there was a significant interaction between the error rates and the presentation order of the tasks ($F_{6,288} = 2.35$, $p = .031$, partial $\eta^2 = .047$), indicating that the pattern of performance effects was different for the participants who had performed the task in the very beginning of each sound condition (Group 1) and for those who had performed it at the end of each sound condition (Group 2). Keeping in mind the practical aim of the study, the paired comparisons were only conducted for the latter group ($N = 26$) as their results can be considered as more reliable due to the longer exposure time preceding the task.

Compared to silence, the error rates were significantly increased in speech ($t(25) = 2.8$, $p = .02$, one-tailed, $d = .43$) and in three masked speech conditions, namely the conditions with vocal music ($t(25) = 3.9$, $p = .011$, one-tailed, $d = .55$), ventilation noise ($t(25) = 3.5$, $p = .011$, one-tailed, $d = .40$) and filtered pink noise ($t(25) = 3.0$, $p = .018$, one-tailed, $d = .32$). Compared to speech, a decrease in error rates was only observed when speech was masked with spring water sound ($t(25) = 2.5$, $p = .035$, one-tailed, $d = .51$). Paired comparisons of the five masked speech conditions indicated that spring water sound as a masking sound improved performance in comparison to vocal music ($t(25) = 3.3$, $p = .021$, two-tailed, $d = .64$) and ventilation noise ($t(25) = 2.7$, $p = .039$, two-tailed, $d = .48$).

In the creative thinking task, a marginal effect of sound condition was observed for the subjects' ideational originality ($F_{6,306} = 1.93$, $p = .075$, partial $\eta^2 = .037$, Figure 5). The effect was not moderated by the presentation order of tasks as in the serial recall task ($p = .70$ for interaction). Ideational fluency (i.e. number of ideas) did not differ between the sound conditions ($F_r = 5.98$, $df = 6$, $p = .43$).

The performance in the proofreading task was not affected by the sound conditions as the analysis showed neither a main effect for the sound condition ($F_{6,306} = 1.51$, $p = .17$) nor an interaction between the sound condition and the detection of different error types ($F_{5,237} = 1.09$, $p = .37$) (Table III). A separate analysis of the reading speed showed no main effect of the sound condition ($F_r = 6.53$, $df = 6$, $p = .37$).

3.2. Subjective measures

There was a significant main effect of sound condition on acoustic satisfaction ($F_{6,318} = 55$, $p < .001$, partial $\eta^2 = .51$, Table II, Figure 6). Paired comparisons revealed that speech ($p < .001$, one-tailed, $d = 2.8$) and all the masked speech conditions ($p < .001$ for all, one-tailed, $d = 2.0$–2.5) were experienced as significantly less satisfying than silence, as hypothesized. Acoustic satisfaction was higher in all masked speech conditions compared to speech ($p = .008$ for all, one-tailed, $d = 0.35$–0.81). Paired comparisons between masked speech conditions revealed that using spring water sound as a speech masker resulted in higher acoustic satisfaction than using vocal music ($t(53) = 3.8$, $p < .001$, two-tailed, $d = .46$). Acoustic satisfaction was also higher in the filtered pink noise condition than in either musical conditions, namely vocal music ($t(53) = 3.5$, $p = .002$, two-tailed, $d = .45$) and instrumental music ($t(53) = 2.3$, $p = .042$, two-tailed, $d = .31$).

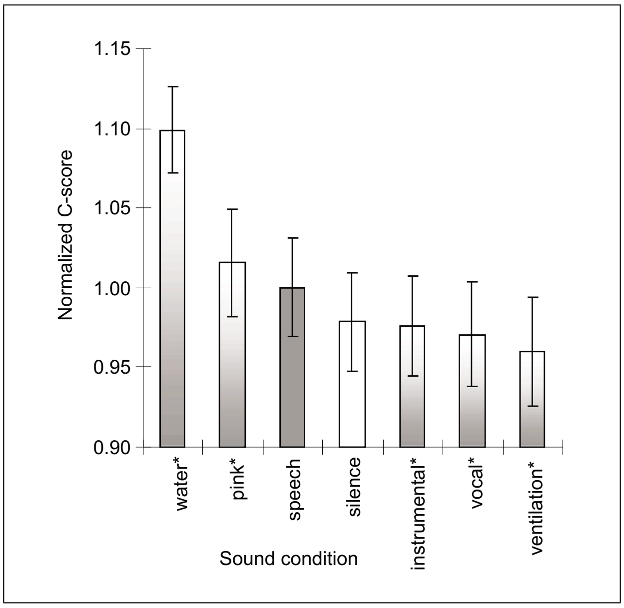

Figure 5. The normalized C-score of the ideational originality in the creative thinking task. Means and standard errors in the seven sound conditions. Values above 1.00 indicate better ideational originality compared to the mean ideational originality in seven sound conditions. The conditions with an asterisk (*) are masked speech conditions in which the sound was presented together with speech.

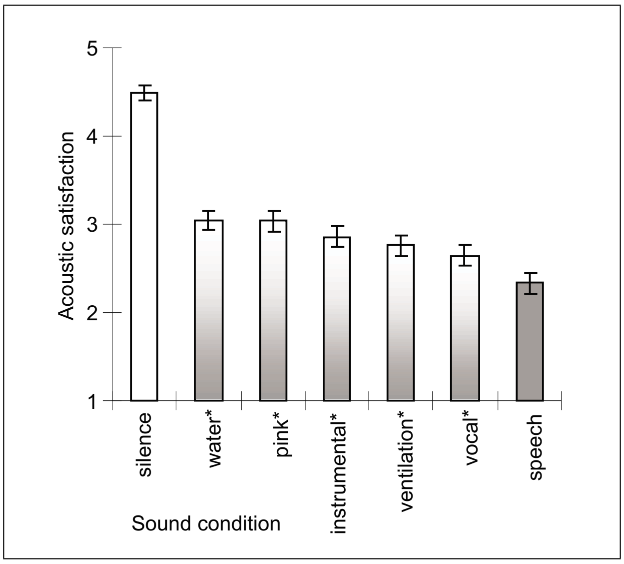

Figure 6. Acoustic satisfaction in the seven sound conditions. Means and standard errors. The conditions with an asterisk (*) are masked speech conditions in which the sound was presented together with speech.

The analysis of subjective workload revealed a significant main effect of sound condition ($F_{6,318} = 41$, $p < .001$,

Table III. Means and standard deviations (in brackets) for total number of errors (max. 60), non-contextual errors (max. 30), and contextual errors (max. 30) and the number of lines read (max. 140 lines) in the proofreading task in different sound conditions. The conditions with an asterisk (*) are masked speech conditions in which the sound was presented together with speech.

Sound condition	Total	Non-contextual errors	Contextual errors	Lines read
silence	38.6 (9.9)	19.4 (5.2)	19.2 (5.9)	118 (18)
speech	36.9 (10.2)	18.6 (5.0)	18.4 (6.3)	117 (22)
pink*	39.1 (9.5)	20.0 (5.1)	19.2 (6.1)	121 (18)
ventilation*	37.2 (9.6)	18.1 (5.1)	19.3 (5.8)	118 (21)
instrumental*	38.5 (8.9)	19.6 (4.4)	18.9 (5.8)	120 (19)
vocal*	37.9 (9.3)	19.2 (4.8)	18.9 (6.6)	119 (22)
water*	37.6 (9.6)	18.9 (5.3)	18.7 (5.2)	118 (21)

partial $\eta^2 = .43$), with a similar pattern of results as acoustic satisfaction. Ratings of subjective workload were significantly lower in silence than in the other sound conditions ($p < .001$ for all, one-tailed, $d = 1.3$–2.0). Compared to speech, subjective workload was lower when speech was masked with filtered pink noise ($p < .001$, one-tailed, $d = .60$), spring water sound ($p = .002$, one-tailed, $d = .43$), ventilation noise ($p = .007$, one-tailed, $d = .46$) and instrumental music ($p = .041$, one-tailed, $d = .27$). It is notable that subjective workload did not differ between speech and the vocal music condition ($p = .34$, one-tailed), implying that even though the condition with vocal music was experienced as more pleasant than speech, it was not experienced as an easier environment for performing tasks. Paired comparisons of masked speech conditions further emphasized the negative experience of vocal music as it differed significantly from all other masked speech conditions ($p = .015$, two-tailed, $d = 0.33$–0.65). The use of instrumental music also resulted in higher subjective workload than the use of filtered pink noise ($p = .026$, two-tailed, $d = .31$).

Additional evidence for the harmful effects of the two music conditions was obtained from the subjective ratings of the different noise sources in each condition (Figure 7). The subjects were more disturbed by vocal music as a masking sound than by any of the other four masking sounds ($p < .001$). Similarly, instrumental music was experienced as more disturbing as a masking sound than the three continuous and temporally almost constant masking sounds, namely spring water sound, filtered pink noise and ventilation noise ($p < .001$). The disturbance caused by the continuous masking sounds was minor in relation to the disturbance of simultaneous speech, as would be required of an optimal masking sound.

No additional information was gained from the analysis of arousal as it was not significantly affected by different sound conditions ($p = .113$). The presentation order of the tasks had no effect on any of the subjective measures.

4. Discussion

4.1. Main findings

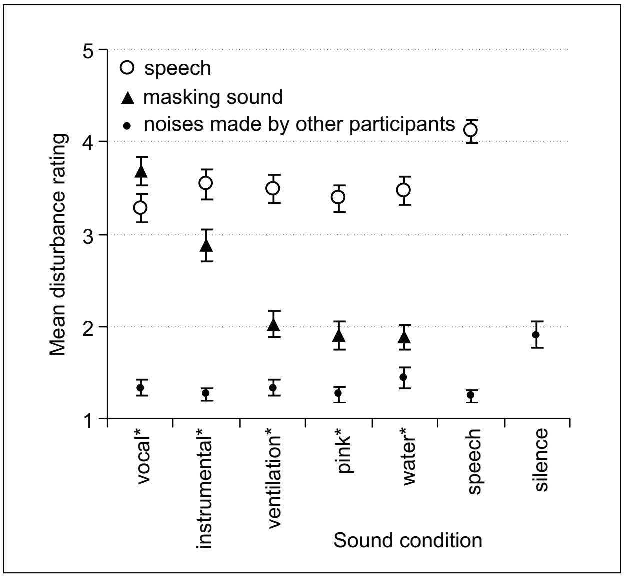

Figure 7. The subjective disturbance caused by the sounds which existed in the laboratory room during the sound conditions (1 = not at all, 5 = very much). Means and standard errors. The conditions with an asterisk (*) are masked speech conditions in which the sound was presented together with speech.

The goals of this study were to show that a reduction of the intelligibility of irrelevant background speech improves cognitive performance and subjective comfort and, more importantly, to compare five masking sounds to determine whether the effect of speech intelligibility depends on the masking sound type. The findings of the study showed that unmasked and highly intelligible speech had detrimental effects on both acoustic comfort and cognitive performance (question 1). When masking sound was used to decrease speech intelligibility, i.e. the detrimental effects of speech, a general improvement of acoustic satisfaction was observed with all masking sounds (research question 2). However, only spring water sound generated a significant improvement in objectively measured performance compared to speech (question 2). Of the different masking sounds, spring water sound had most benefits in terms of both subjective and objective indicators, whereas vocal music was as detrimental to performance as speech (question 3). The findings are roughly summarized in Table IV.

To address the first two research questions in more detail, the results of the subjective measures strongly support the hypotheses. Silence resulted in the highest acoustic satisfaction and the lowest subjective workload when

Table IV. A summary of the conclusions regarding research questions Q1–Q3 for the five masked speech conditions. A "+" sign indicates a statistically significant improvement in comparison to the reference sound condition. A "−" sign indicates a statistically significant decrement in comparison to the reference sound condition. Empty cells indicate that statistically significant differences were not found.

	water	pink	ventilation	instrumental	vocal
Q1: Comparison to silence					
Performance in serial recall task		−	−		−
Acoustic satisfaction	−	−	−	−	−
Subjective workload	−	−	−	−	−
Q2: Comparison to speech					
Performance in serial recall task	+				
Acoustic satisfaction	+	+	+	+	+
Subjective workload	+	+	+	+	
Q3: Comparison between masking sounds					
Performance in serial recall task	+		−		−
Acoustic satisfaction	+	+		−	−
Subjective workload	+	+	+	−	−
Disturbance of masking sound itself	+	+	+	−	−
Overall ranking order	1	2	3	4	5

compared to both unmasked and masked speech (question 1). The observed improvement in subjective measures between unmasked and masked speech (question 2) is in line with earlier findings that subjective comfort increases with decreasing speech intelligibility (e.g. [18, 20, 26, 36]). The only exception of this relation was vocal music which improved acoustic satisfaction but did not decrease subjective workload, whereas the other four masking sounds improved both.

The results of the performance tasks were not as clear in relation to the first two hypotheses as the subjective results. Only a marginal effect was observed in creative thinking task, and no effect in proofreading. However, sound conditions had an effect on the serial recall task which is a classical task for investigating performance effects of speech. A rise in error rates was observed between silence and speech conditions, repeating an irrelevant sound effect (question 1). The same effect occurred when speech was masked with filtered pink noise, ventilation sound and vocal music but, contrary to expectations, not when spring water sound and instrumental music were used (question 1). When the effect of masked speech was examined in the opposite direction, namely in relation to speech condition (question 2), an improvement in performance was only obtained by masking speech with spring water sound.

The third research question examined differences between the masking sounds with no specific hypotheses. In line with the findings above, the serial recall results distinguished the spring water sound as a more effective masking sound than the other four. Similar trend was also observed in the creative thinking task (Figure 5). Whereas the performance results highlighted the positive effects of spring water sound, the differences in acoustic satisfaction and subjective workload emphasized the adverse effects of vocal music in relation to the other masked speech conditions. Vocal music was rated as disturbing as the background speech (Figure 7). There was less evidence for the effects of the remaining three masking sounds. Filtered pink noise seemed to position close to spring water sound in terms of subjective perceptions, whereas instrumental music was associated with contradictory evidence. Ventilation noise was not experienced as disturbing, but there were no significant indications of its positive effects either.

4.2. Effects of masked speech on performance

Overall, the serial recall results of this study are not compatible with either the changing-state hypothesis [28] or the STI model of Hongisto [13]. The changing-state hypothesis would predict that the performance decrement in serial recall depends on the variability of the background sound (see Introduction), whereas the STI model would predict that it depends on speech intelligibility, as determined by STI. The variability of the sound environments was not determined by sophisticated sound quality descriptors, such as fluctuation strength [58], but it is obvious that spring water sound, ventilation sound and filtered pink noise were temporally more continuous sounds than the two musical maskers. Therefore, it is unexpected that speech masked with spring water sound and instrumental music did not produce a decline in performance but speech masked with ventilation sound and filtered pink noise did. On the other hand, the STI model would predict no differences between the masked speech conditions because the STI values were equal (Table I). Thus, the pattern of results cannot be explained by acoustic variability or speech intelligibility alone.

One explanation for the different effects of instrumental and vocal music can be the verbal content of vocal music. Both music types contained instruments which are effective speech maskers but vocal music added another source of verbal material which goes against the underlying aim of speech masking. Two simultaneous voices are not enough for creating speech babble that would sufficiently reduce speech intelligibility and the associated distraction [42]. This explanation is also supported by previous findings that vocal music is more detrimental to cognitive performance than instrumental music when tested independently without simultaneous speech [19, 24]. Furthermore, the different effects of the two types of music may imply that the semantic content of the background sound contributes to the irrelevant sound effect. This would contradict the currently dominant view in the changing-state paradigm (e.g. [28]).

Why did the spring water sound outperform the other masking sounds in the serial recall task? The subjective perceptions do not provide additional information. One explanation may be the measurement method of the STI (Table I). The method assumes a constant background noise. Therefore, the speech-to-noise was determined from the equivalent levels of speech and masking sounds. That is, the temporal variations of masking were ignored. This was also expected in the STI model of Hongisto [13]. The spring water sound contained rapid level modulations which may coincide with the fastest level modulations of speech at 8–15 Hz. It may be that a more advanced method for determining the STI, which would consider the modulations of background noise with respect to speech, would have resulted in a lower STI value than the static STI method did. In other words, the condition with spring water as a masking sound may have resulted in lower subjective speech intelligibility than those with more continuous masking sounds, e.g., filtered pink noise and ventilation noise. In the future, more emphasis should be placed on the analysis of temporal characteristics of masking sound instead of only measuring the equivalent sound pressure levels. It is expected that if the modulations of masking sound coincide profitably with the modulations of speech, the masking efficiency may increase significantly compared to a masking sound which has no modulations but equal sound pressure level and spectrum. We suggest that the STI model of Hongisto [13] is most valid for masking sounds that are temporally constant, such as pink noise or ventilation noise.

An alternative, or additional, explanation for the effectiveness of the spring water sound might be drawn from the largely documented psychological benefits of natural environments to which this sound obviously belongs. Natural environments have been shown to have restorative effects on cognitive performance (e.g. [59, 60]) and physiological stress recovery (e.g. [61]). To our knowledge, the role of the auditory component in a restorative process has only been tested in one very recent experimental study [62]. Restoration refers to the recovery from an emotionally or cognitively demanding situation, but if auditory experiences were to play a role in restoration, it seems possible that certain auditory stimuli might also protect the cognitive resources during cognitively demanding situations. Currently, this aspect of speech masking sounds has not been addressed in the literature at all.

Proofreading performance was also measured but it was not affected by the sound condition. Similar results have recently been reported by Haka *et al.* [20] and Smith-Jackson and Klein [63]. On the other hand, effects of highly intelligible speech on proofreading have also been reported [18, 28]. The effect of background speech on semantically-oriented tasks, such as proofreading or reading comprehension, seems to be generally smaller than the irrelevant sound effect observed in serial recall [13, 28]. The lack of significant findings could therefore be explained by the narrow range of STI values used in this study (STI 0.00–0.62) and a slightly shorter, and less sensitive, task format than the one used by Jones *et al.* [28]. The observation that people seem to be able to resist the distracting effects of speech better in the proofreading task than in serial recall may be explained by the differences in the cognitive demands. The serial recall task assesses the short-term memory store which has very limited capacity and is known to be susceptible to acoustic distraction. However, the effect of speech on proofreading may not be mediated by short-term memory [28]. Proofreading is a complex task which requires a wider range of cognitive skills, such as visual word recognition, working memory and semantic processing. Due to the higher complexity, proofreading task may also allow more room for compensating for the distraction by switching between different performance strategies.

4.3. Music as speech masker for office work

The only contradictory evidence in the present study involved the use of instrumental music (Table IV). The absence of an irrelevant sound effect in serial recall when speech was masked with instrumental music gives some support for its use as a masking sound. In terms of acoustic satisfaction and subjective workload, instrumental music seemed to work as a speech masker but its effects on these subjective indicators were significantly smaller than those obtained by using filtered pink noise. Also, instrumental music was rated significantly more disturbing as a masking sound than filtered pink noise, ventilation noise and spring water sound (Figure 7). The applicability of instrumental music to open-plan offices should be viewed with some reservations, even though it did not have as disturbing effect on performance as vocal music did.

But why do some workers voluntarily listen to music even though it has such negative effects? The uncontrollability of sound is in general associated with the negative experience of noise [64]. Thus, choosing to listen to music could be explained by the associated increase in the experienced control over the acoustic environment. The present results also imply that individual differences play a role in whether music is liked or disliked as a masking sound, as the two music conditions tended to result in the largest standard deviations in the individual items in acoustic satisfaction (Table II). A recent experimental study has shown that the effect of music on performance differs between introverts and extroverts, and the latter also report a strong likelihood of studying with background music [65]. A field study by Nemecek [39] has found that the disturbance of background music increases with increasing cognitive or mental demands of the job. Taking the large variations between individuals and different work tasks into account, masking sound technologies that utilize music, especially vocal music, may not be appropriate for open-plan offices in general.

4.4. Subjective vs. objective measures

Why were the differences hypothesized in research questions 1–3 more strongly observed in the subjective measures than in the objectively measured performance? The finding that subjective comfort is more sensitive to differences between acoustic environments than performance measures has also been observed in previous studies (e.g. [18, 20, 26]. One explanation for this is offered by the enhanced effort hypothesis [26]. The subjective awareness of disturbance is assumed to lead individuals to invest more effort in the task to compensate for the anticipated performance loss. Such efforts could moderate the effect of environmental distraction on objectively measured performance. In the current study, such compensatory efforts should be reflected in the variable of subjective workload, reflecting the psychological and emotional costs of trying to maintain performance in the presence of auditory interference. In the working life, such a performance-cost trade-off would very likely have implications for the overall work performance of an individual. In other words, the decrease in subjective workload in masked speech conditions compared to unmasked speech may anticipate performance improvement in similar conditions over time even though this study observed only a modest improvement.

Of the different subjective measures, arousal was the only one not affected by sound conditions. It was included in the study to gain supplementary information on the potential harmful effects of masking sounds as noise-related increases in arousal may impair performance on complex tasks [37]. The lack of differences in self-reported arousal implies that the level range used in this study (37–48 dB, Table I) does not lead to changes in arousal. Another possibility is that the subjective measurement was too rough for capturing relevant changes in arousal. Physiological measures of stress response could be useful in detecting both beneficial and harmful effects of different sounds in future experiments. However, physiological measures may be more ambiguous and less sensitive than questionnaire measures in this area of research [62], and should therefore only be used for complementary information.

4.5. Methodological issues

The present findings have some methodological implications for research on masking sounds and speech sounds in general. First of all, the unexpected interaction between serial recall performance and the task order shows that the exposure time to the sound environment needs to be considered, particularly when using more variable sounds. As Figure 4 shows, the pattern of serial recall performance differed between those participants that had a 15-to-20-minutes exposure to the sound condition before the task and those that did not. We interpret this finding as a habituation effect that concerns particularly the spring water sound and, to some extent, instrumental music. That is, these sounds were initially distracting but were quickly accustomed to. Thus, including habituation periods to experiments and conducting longer test sessions would increase the reliability of the conclusions on sound effects.

The limitations of this experiment concern the typical restrictions in generalizing findings from laboratory environments to real work environments. The cognitive demands of real office work can vary a lot, and there are many other factors that may moderate the effects of the physical work environment. Different masking sounds should also be investigated in office environments where the technical implementation is carefully made, see e.g., Hongisto [46].

In addition, the seven acoustic conditions and several research questions made the research design complex, particularly as the effect of task order had to be incorporated into the analysis. These factors may have weakened the power of the study to reveal expected differences.

4.6. Practical considerations

Speech privacy should be high in open-plan offices where individual work with high concentration demands or private conversations are emphasized. In principle, such jobs belong to private office rooms but because open-plan offices are favored in most organizations, careful speech privacy design is necessary for compensating the absence of full-height partition walls. In such open-plan offices, all room acoustic measures moderating the level of speech should be considered simultaneously, i.e. high absorption in room boundaries (ceiling, walls, floor, furniture, hanging absorbers), high screens and high degree of enclosure around the workstation. In addition, a speech masking system is strongly recommended.

Carefully designed speech masking can reduce the negative effects of irrelevant speech on cognitive work performance and acoustic satisfaction. Vocal or instrumental music are not recommended to be used as a speech masking sound. The most common masking sounds, filtered pink noise and ventilation, are acceptable and also very practical alternatives. More advanced alternatives could be wide-band natural sounds, like the spring water sound, which was the most beneficial masking sound in this experiment.

The speech masking system should be individually designed for each workplace. It should be considered only when the initial level of A-weighted background noise level in the occupied room is far below 40 dB. There are many alternative ways to install the masking loudspeakers. The most appropriate technical solution depends on, e.g., the acoustical properties, furniture, height and the geometry of the room. The A-weighted level of masking sound should not exceed 45 dB to enable normal conversations but not below 40 dB in order to achieve effective speech masking.

Acknowledgement

This research was carried out as a part of a national research programme MAKSI (Perceived and Modeled Indoor Environment) funded by Tekes, the National Technology Agency, the Finnish Institute of Occupational Health and various companies.

References

[1] E. M. de Croon, J. K. Sluiter, P. P. Kuijer, M. H. Frings-Dresen: The effect of office concepts on worker health and performance: a systematic review of the literature. Ergonomics **48** (2005) 119–134.

[2] J. Pejtersen, L. Allermann, T. S. Kristensen, O. M. Poulsen: Indoor climate, psychosocial work environment and symptoms in open-plan offices. Indoor Air **16** (2006) 392–401.

[3] A. Haapakangas, R. Helenius, E. Keskinen, V. Hongisto: Perceived acoustic environment, work performance and well-being – survey results from Finnish offices. 9th International Congress on noise as a Public Health Problem (ICBEN), Mashantucket, Connecticut, USA, 2008, July 21-25.

[4] 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.

[5] C. Bodin Danielson, L. Bodin: Difference in satisfaction with office environment among employees in different office types. Journal of Architectural and Planning Research **26** (2009) 241–257.

[6] K. L. Jensen, E. Arens, L. Zagreus: Acoustical quality in office workstations as assessed by occupant surveys. Proceedings of Indoor Air, 10th international conference on indoor air quality and climate, Beijing, China, 4-9 September 2005, 2401–2405.

[7] E. Sundstrom, J. P. Town, R. W. Rice, D. P. Osborn, M. Brill: Office noise, satisfaction, and performance. Environment and Behaviour **26** (1994) 195–222.

[8] J. S. Bradley: The acoustical design of conventional open plan offices. Canadian Acoustics **31** (2003) 23–31.

[9] P. Virjonen, J. Keränen, R. Helenius, J. Hakala, V. Hongisto: Speech privacy between neighbouring workstations in an open office. A laboratory study. Acta Acustica united with Acustica **93** (2007) 771–782.

[10] P. Virjonen, J. Keränen, V. Hongisto: Determination of acoustical conditions in open-plan offices. Proposal for new measurement method and target values. Acta Acustica united with Acustica **95** (2009) 279–290.

[11] H. C. Hardy: A guide to office acoustics. Architectural Record, February 1957.

[12] R. A. Waller: Office acoustics. Effect of background noise. Applied Acoustics **2** (1969) 121–130.

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

[14] H. A. Colle, A. Welsh: Acoustic masking in primary memory. Journal of Verbal Learning and Verbal Behaviour **15** (1976) 17–31.

[15] D. M. Jones, N. Morris: Irrelevant speech and cognition. – In: Handbook of Human Performance, 1. D. M. Jones, A. P. Smith (eds.). Academic Press, London, 1992, 29–53.

[16] C. P. Beaman, D. M. Jones: The role of serial order in the irrelevant speech effect: tests of the changing-state hypothesis. Journal of Experimental Psychology: Learning, Memory and Cognition **23** (1997) 459–471.

[17] D. M. Jones, C. Miles, J. Page: Disruption of proof-reading by irrelevant speech: effects of attention arousal or memory? Applied Cognitive Psychology **4** (1990) 89–108.

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

[19] R. C. Martin, M. S. Wolgalter, J. G. Forlano: Reading comprehension in the presence of unattended speech and music. Journal of Memory and Language **27** (1988) 382–398.

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

[21] U. Landström, L. Söderberg, A. Kjellberg, B. Nordström: Annoyance and performance effects of nearby speech. Acta Acustica United with Acustica **88** (2002) 549–553.

[22] W. J. Macken, S. Tremblay, D. Alford, D. M. Jones: Attentional selectivity in short-term memory: Similarity of process, not similarity of content, determines disruption. International Journal of Psychology **34** (1999) 322–327.

[23] D. M. Jones, C. Madden, C. Miles: Privileged access by irrelevant speech to short-term memory: The role of changing state. The Quarterly Journal of Experimental Psychology **44A** (1992) 645–669.

[24] P. Salamé, A. Baddeley: Differential effects of noise and speech on short-term memory. Noise as a Public Health Problem (1983); Proceedings of the Fourth International Congress, Turin, Italy, June 21-25, 2008, 751–758.

[25] H. A. Colle: Auditory encoding in visual short-term recall: Effects of noise intensity and spatial location. Journal of Verbal Learning and Verbal Behavior **19** (1980) 722–735.

[26] 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.

[27] S. Tremblay, D. M. Jones: Change of intensity fails to produce irrelevant sound effect: Implications for the representation of unattended sound. Journal of Experimental Psychology: Human Perception and Performance **25** (1999) 1005–1015.

[28] D. M. Jones, W. J. Macken: Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. Journal of Experimental Psychology: Learning, Memory and Cognition **19** (1993) 369–381.

[29] H. Nittono: Background instrumental music and serial recall. Perceptual and Motor Skills **84** (1997) 1307–1313.

[30] N. Morris, D. M. Jones, A. J. Quayle: Memory disruption by background speech and singing. – In: Contemporary ergonomics. E. D. Megan (ed.). Taylor and Francis, London, 1989, 494–499.

[31] S. J. Schlittmeier, J. Hellbrück, M. Klatte: Does irrelevant music cause an irrelevant sound effect for auditory items? European Journal of Cognitive Psychology **20** (2008) 252–271.

[32] D. C. LeCompte, C. B. Neely, J. R. Wilson: Irrelevant speech and irrelevant tones: The relative importance of speech to the irrelevant speech effect. Journal of Experimental Psychology: Learning, Memory, and Cognition **23** (1997) 472483.

[33] A. Buchner, K. Rothermund, D. Wentura, B. Mehl: Valence of distractor words increases the effects of irrelevant speech on serial recall. Memory and Cognition **32** (2004) 722–731.

[34] W. Ellermeier, J. Hellbrück: Is level irrelevant in 'irrelevant speech'? Effects of loudness, signal-to-noise ratio and binaural masking. Journal of Experimental Psychology: Learning, Memory and Cognition **24** (1998) 1406–1414.

[35] A. Liebl, H. Drotleff, R. Wack, M. Leistner: The influence of speech intelligibility on performance and acoustic comfort. Proceedings of 39th International Congress on Noise Control Engineering, Internoise, Lisbon, Portugal, 13-16 June 2010.

[36] J. Veitch, J. Bradley, L. Legault, S. Norcross, J. Svec: Masking speech in open-plan offices with filtered pink noise noise: Noise level and spectral composition effects on acoustic satisfaction. Institute for Research in Construction, Internal Report IRC-846, 2002, Ottawa, Canada, 2002.

[37] L. Loewen, P. Suedfield: Cognitive and arousal effects of masking office noise. Environment and Behaviour **24** (1992) 381–395.

[38] E. Sundstrom: Work places: The psychology of the physical environment in offices and factories. Cambridge University Press, New York, 1986.

[39] J. Nemecek: Music during office work. – In: Ergonomics and Health in Modern Offices. E. Grandjean (ed.). Taylor and Francis, London, 1984, 64–69.

[40] G. Oldham, A. Cummings, L. Mischel, J. Schmidtke, J. Zhou: Listen while you work? quasi-experimental relations between personal-stereo headset use and employee work responses. Journal of Applied Psychology **80** (1995) 547–564.

[41] S. J. Schlittmeier, J. Hellbrück: Background music as noise abatement in open-plan offices: A laboratory study on performance effects and subjective preferences. Applied Cognitive Psychology **23** (2009) 684–697.

[42] D. M. Jones, W. J. Macken: Auditory babble and cognitive efficiency. Role of number of voices and their location. Journal of Experimental Psychology - Applied **1** (1995) 216–226.

[43] WHO: Noise and health. – In: Local authorities, health and environment briefing pamphlet series, 35. X. Bonnefoy, B. Berglund, M. C. (eds.). World Health Organization, Copenhagen, 2000.

[44] M. Navai, J. Veitch: Acoustic satisfaction in open-plan offices: Review and recommendations. Institute for Research in Construction, Research Report IRC-RR-151, 2003, Ottawa, Canada, 2003.

[45] R. Helenius, V. Hongisto: The effects of acoustical improvement of an open-plan office on workers. Proceedings of Inter-Noise 2004, Prague, Czech Republic, 21-25 August 2004, Paper 674.

[46] V. Hongisto: Effects of sound masking on workers. A case study in a landscape office. 9th International Congress on noise as a Public Health Problem (ICBEN), Mashantucket, Connecticut, USA, July 21-25, 2008.

[47] IEC 60268-16: Sound system equipment. Part 16: Objective rating of speech intelligibility by speech transmission index. International Electrotechnical Commission, Geneva, Switzerland, 1998.

[48] V. Hongisto, J. Keränen, P. Larm: Simple model for the acoustical design of open-plan offices. Acta Acustica united with Acustica **90** (2004) 481–495.

[49] L. A. Reyner, J. A. Horne: Falling asleep while driving: Are drivers aware of prior sleepiness? International Journal of Legal Medicine **111** (1998) 12–23.

[50] T. Åkerstedt, M. Gillberg: Subjective and objective sleepiness in the active individual. International Journal of Neuroscience **52** (1990) 29–37.

[51] D. P. Wyon: Creative thinking as the dependent variable in six environmental experiments: A review. Proceedings of the 7th International Conference on Indoor Air Quality and Climate – Indoor Air '96, Nagoya, Japan, 1996, 1, 419–422.

[52] J. P. Guildford, P. R. Christensen, P. R. Merrifield, R. C. Wilson: Alternate uses: Manual of instructions and interpretation. Sheridan Psychological Services, Orange, California, USA, 1978.

[53] K. Ahonen, J. Rosenqvist, P. Valotie: Blow your mind. locating new Finnish moving pictures (in Finnish). Kirja-Aurora, Turku, Finland, 2003.

[54] N. D. Weinstein: Short notes, noise and intellectual performance: A confirmation and extension. Journal of Applied Psychology **62** (1977) 104–107.

[55] Y. Benjamini, Y. Hochberg: Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B **57** (1995) 289–300.

[56] Y. Benjamini, D. Yekutieli: The control of the false discovery rate in multiple testing under dependency. The Annals of Statistics **29** (2001) 1165–1188.

[57] J. Cohen: Statistical power analysis for the behavioral sciences. 2nd ed. Lawrence Erlbaum Associates Inc. Publishers, Hillsdale, New Jersey, USA, 1988.

[58] S. J. Schlittmeier, T. Weißgerber, S. Kerber, H. Fastl, J. Hellbrück: An algorithm modelling the irrelevant sound effect (ISE). Proceedings of the Acoustics'08. 5th Forum Acusticum of the European Acoustics Association, Paris, France, June 29–July 4, 2008, 1197–1201.

[59] T. Hartig, M. Mang, G. W. Evans: Restorative effects of natural environment experiences. Environment and Behavior **23** (1991) 3–26.

[60] R. Berto: Exposure to restorative environments helps restore attentional capacity. Journal of Environmental Psychology **25** (2005) 249–259.

[61] R. S. Ulrich, R. F. Simons, B. D. Losito, E. Fiorito, M. A. Miles, M. Zelson: Stress recovery during exposure to natural and urban environments. Journal of Environmental Psychology **11** (1991) 201–230.

[62] H. Jahncke, S. Hygge, N. Halin, A.-M. Green, K. Dimberg: An experiment on noise and cognition in a simulated open-plan office. Proceedings of Internoise 2010, The 39th International Congress and Exposition on Noise Control Engineering, Lisbon, Portugal, 13-16 June 2010.

[63] T. L. Smith-Jackson, K. W. Klein: Open-plan offices: Task performance and mental workload. Journal of Environmental Psychology **29** (2009) 279–289.

[64] U. Sailer, M. Hassenzahl: Assessing noise annoyance: an improvement oriented approach. Ergonomics **43** (2000) 1920–1938.

[65] T. Chamorro-Premuzic, V. Swami, A. Terrado, A. Furnham: The effects of background auditory interference and extraversion on creative and cognitive task performance. International Journal of Psychological Studies **1** (2009) 2–9.

[66] S. Banbury, D. Berry: Habituation and dishabituation to speech and office noise. Journal of Experimental Psychology: Applied **3** (1997) 181–195.

[67] P. Larm, V. Hongisto: Experimental comparison between speech transmission index, rapid speech transmission index, and speech intelligibility index. Journal of the Acoustical Society of America **119** (2006) 1106–1117.