Unmasking the effects of masking on performance: The potential of multiple-voice masking in the office environment

Citation for the original published paper (version of record):
Keus Van De Poll, M., Carlsson, J., Marsh, J. et al (2015). Unmasking the effects of masking on performance: The potential of multiple-voice masking in the office environment. Journal of the Acoustical Society of America, 138(2): 807-816.
http://dx.doi.org/10.1121/1.4926904

The Journal of the Acoustical Society of America

AUGUST 17 2015

Unmasking the effects of masking on performance: The potential of multiple-voice masking in the office environment

Marijke Keus van de Poll; Johannes Carlsson; John E. Marsh; Robert Ljung; Johan Odelius; Sabine J. Schlittmeier; Gunilla Sundin; Patrik Sörqvist

J. Acoust. Soc. Am. 138, 807–816 (2015)
https://doi.org/10.1121/1.4926904

Unmasking the effects of masking on performance: The potential of multiple-voice masking in the office environment

Marijke Keus van de Poll
Department of Building, Energy and Environmental Engineering, University of Gävle, Kungsbäcksvägen 47, SE-801 76 Gävle, Sweden

Johannes Carlsson
Division of Applied Acoustics, Department of Civil and Environmental Engineering, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden

John E. Marsh
School of Psychology, University of Central Lancashire (UCLan), DB 115, Darwin Building, Preston, PR1 2HE, United Kingdom

Robert Ljung
Department of Building, Energy and Environmental Engineering, University of Gävle, Kungsbäcksvägen 47, SE-801 76 Gävle, Sweden

Johan Odelius
Department of Civil, Environmental and Natural Resources Engineering, Luleå University of Technology, Laboratorievägen 14, SE-971 87 Luleå, Sweden

Sabine J. Schlittmeier
Work, Environmental and Health Psychology, Catholic University of Eichstätt-Ingolstadt, Ostenstraße 25, 85072 Eichstätt, Germany

Gunilla Sundin
Akustikon Team in Norconsult AB, Hantverkargatan 5, SE-112 21 Stockholm, Sweden

Patrik Sörqvist[a)]
Department of Building, Energy and Environmental Engineering, University of Gävle, Kungsbäcksvägen 47, SE-801 76 Gävle, Sweden

(Received 5 December 2014; revised 10 May 2015; accepted 2 July 2015; published online 17 August 2015)

Broadband noise is often used as a masking sound to combat the negative consequences of background speech on performance in open-plan offices. As office workers generally dislike broadband noise, it is important to find alternatives that are more appreciated while being at least not less effective. The purpose of experiment 1 was to compare broadband noise with two alternatives—multiple voices and water waves—in the context of a serial short-term memory task. A single voice impaired memory in comparison with silence, but when the single voice was masked with multiple voices, performance was on level with silence. Experiment 2 explored the benefits of multiple-voice masking in more detail (by comparing one voice, three voices, five voices, and seven voices) in the context of word processed writing (arguably a more office-relevant task). Performance (i.e., writing fluency) increased linearly from worst performance in the one-voice condition to best performance in the seven-voice condition. Psychological mechanisms underpinning these effects are discussed.
[http://dx.doi.org/10.1121/1.4926904]

[JFC] Pages: 807–816

I. INTRODUCTION

The negative effects of open-plan office noise (Jahncke *et al.*, 2011), and especially of background speech (Marsh and Jones, 2010), on cognitive performance are well documented. For example, background speech tends to impair tasks like proofreading (Halin *et al.*, 2013), reading comprehension (Halin *et al.*, 2014; Sörqvist *et al.*, 2010; Perham and Currie, 2014), and writing (Keus van de Poll *et al.*, 2014; Sörqvist *et al.*, 2012) that are presumably representative of the type of task typically undertaken in office environments. Many companies have introduced masking sound in the working environment to combat the potential negative impact of background speech on such cognitive tasks. Typically, the masking sounds used within the office setting are continuous broadband noises (or so-called aperiodic sounds that have no clear cycle of repetition and thus lack a clear tonal quality)

[a)]Electronic mail: patrik.sorqvist@hig.se

such as pink noise, but they are not generally appreciated by the office workers (Haapakangas *et al.*, 2011; Schlittmeier and Hellbrück, 2009). The purpose of the present study was to explore the effectiveness of alternatives to pink noise as masking sound to identify more appropriate maskers for performance protection against disruption by background speech. In experiment 1, the aim was to explore the effects of several alternative maskers, like multiple voices or water waves, in the context of an often-used memory task, namely, verbal serial recall, since this task has been used in other applied studies on the detrimental impact of background speech and the potential of adding masking sound to reduce it (e.g., Haapakangas *et al.*, 2011; Schlittmeier and Hellbrück, 2009). Experiment 2 delineates the potential benefits of the "winner" in experiment 1—multiple-voice masking—in the context of a more applied task.

A. Why sound disrupts cognitive performance

A changing sound sequence, such as "f m k l c v," is more disruptive than a steady-state sound sequence like "f f f f f f," at least when the task is to recall a sequence of to-be-remembered items in their order of presentation (serial short-term memory). This well-established finding is called the changing-state effect (Jones and Macken, 1993). Several applied implications arise from the fact that more acoustically variable sound is more disruptive than less acoustically variable sound. For example, longer reverberation times are associated with smaller disruption (Beaman and Holt, 2007), although not necessarily within realistic reverberation time intervals (Perham *et al.*, 2007). And background music with less acoustic variation is also less disruptive than music with more acoustic variation (Perham and Sykora, 2012; Schlittmeier *et al.*, 2008). The changing-state effect is the key empirical signature of the irrelevant sound effect: Disruption of serial short-term memory by background sound. According to the interference-by-process view (Jones and Tremblay, 2000), the obligatory, automatic encoding on the order of changing successive items as part of the auditory-streaming, or segmentation, process (Bregman and Rudnicky, 1975), interferes with the deliberate serial ordering process required to rehearse the visually presented items in sequence, thereby disrupting serial recall and giving rise to the irrelevant sound effect. Hence, the factor that makes sound disruptive is its acoustic change.

The way that sound impairs cognitive performance is, however, jointly determined by the properties of the sound and the characteristics of the task (Jones and Tremblay, 2000; Marsh *et al.*, 2009). In the context of reading comprehension (Martin *et al.*, 1988), analytical reasoning (Perham *et al.*, 2005), and writing (Sörqvist *et al.*, 2012), it is not the psycho-acoustical characteristics, but the semantic characteristics of the sound that underpin disruption. This interference-by-process view can also explain the disruptive effect of meaningful background speech on language-based tasks such as writing (Marsh and Jones, 2010). On this view, the involuntary semantic analysis of background speech disrupts performance by interfering with the execution of the deliberate semantic processes that are carried out to fulfil the task requirements (Jones *et al.*, 2012; Marsh *et al.*, 2008; Marsh *et al.*, 2009; Sörqvist *et al.*, 2010; Perham and Currie, 2014). In view of the difference between serial-memory tasks and language-based tasks, a serial short-term memory task was deployed in experiment 1 and a word processed writing task was deployed in experiment 2 of the present study.

B. What happens when the potentially disruptive sound is masked by additional sound?

The acoustic signal of a single voice comprises peaks and troughs (acoustic changes) that can act as the basis of automatic order encoding via segmentation. The addition of continuous noise masks the abrupt pitch and loudness changes in the speech signal. The compound signal (voice and noise) therefore contains fewer cues to segmentation than the single voice signal. One way in which additionally played-back acoustic maskers in office-environments can protect performance is, hence, by reducing the variability of the acoustic signal that reaches the ear—by which the changing-state effect should be attenuated. The reduction of the acoustic variability of the sound stream also reduces speech intelligibility. Another way acoustic maskers can protect performance is, hence, by making it more difficult for the cognitive system to extract meaning from the speech signal—by which disruption that emanates from semantic interference is reduced. Consonant with this, reducing the intelligibility of background speech by superimposing it with a partial masker has been shown to reduce its detrimental effects (e.g., Keus van de Poll *et al.*, 2014).

The speech signal's intelligibility can be quantified in terms of Speech Transmission Index (STI) that varies between 1.00 (perfect intelligibility) to 0.00 (no intelligibility at all). The intelligibility of undesired background speech is also sometimes referred to as "acoustic privacy," quantified as "1 minus STI-value," to emphasize the fact that high privacy values are desirable. In turn, higher STI values are desirable in the context of communication but undesirable in the context of disruption from task-irrelevant background speech. A number of studies have demonstrated that lower STI (or higher privacy) values for task-irrelevant background speech is associated with greater performance (e.g., Ellermeier and Hellbrück, 1998; Haapakangas *et al.*, 2011; Schlittmeier and Hellbrück, 2009). This holds true in visual-verbal serial recall (Ellermeier and Hellbrück, 1998; Ebissou *et al.*, 2013; Schlittmeier and Hellbrück, 2009; Haapakangas *et al.*, 2011) as well as in tasks that require processing of meaning and lexical-based retrieval (Haka *et al.*, 2009; Jahncke *et al.*, 2013; Keus van de Poll *et al.*, 2014; Loewen and Suedfeld, 1992; Venetjoki *et al.*, 2006). Yet, as mentioned above, the reason why background speech is less detrimental when its intelligibility is reduced differs, depending on the task at hand.

Despite the use of aperiodic noise as additionally played-back masking sound, subjective ratings typically indicate low acceptance ratings for continuous noise as a mask (e.g., Schlittmeier and Hellbrück, 2009). In this respect, Haapakangas and co-authors expected enhanced acceptance for less artificial maskers and tested the sound of spring water, instrumental and vocal music as masking sounds (Haapakangas *et al.,* 2011). Superimposing background speech by continuous spring water sound significantly reduced its detrimental impact on serial recall performance. Indeed, performance with the spring water masking sound did not differ significantly from performance in quiet. However, this was only the case for a subgroup of participants and, furthermore, subjective ratings on acoustic satisfaction were still significantly reduced for spring water sounds compared to quiet. Such a discrepancy between performance data and subjective evaluations of masking sounds have also been reported in other studies (e.g., Jiang *et al.,* 2011; Park *et al.*, 2013; Schlittmeier and Hellbrück, 2009; Schlittmeier *et al.*, 2008). These findings underscore the necessity to consider both performance data and subjective ratings when evaluating the effectiveness and appropriateness of a certain sound as a masker in work environments.

Aside from continuous signals like pink noise, discontinuous signals like voices can also function as a speech masker. The effectiveness of this, perhaps less obvious way of masking single voice speech, has been demonstrated a number of times (Ebissou *et al.,* 2013; Hellbrück and Kilcher, 1993; Jones and Macken, 1995; Kittel *et al.*, 2013). For example, Jones and Macken (1995) compared the disruptive effects produced on serial recall of irrelevant sound comprising a single voice with that comprising several simultaneously presented speakers. The disruptive impact was significantly smaller for five or more speakers compared to one single speaker regardless of whether the sound was presented in a language the participant comprehended (see also Hellbrück and Kilcher, 1993; Kittel *et al.,* 2013). This "babble effect" was attributed to the fact that adding voices reduces the degree of perceptible change between successive sounds within the auditory stream, thereby reducing the cues to segmentation and weakening the foundation on which changing state can be brought about.

C. The present study

In view of these considerations, experiment 1 was set up to test the effects of different partial maskers, from natural sounds like water waves to discontinuous speech babble of multiple background talkers, for protecting cognitive performance against the detrimental impact of a single speaker. Here, performance in the often used verbal short-term memory task is supplemented by subjective ratings. The "winning masker" regarding both performance and subjective measures was then selected and explored in more detail within the context of a more applied task in experiment 2, namely, word processed writing.

II. EXPERIMENT 1

The purpose of experiment 1 was to compare performance in five conditions: a single voice condition, three masking conditions (i.e., the single voice masked either by multiple voices, water waves, or pink noise) and a silent control condition. Serial recall was selected as the performance measure and acoustical satisfaction was evaluated by way of a questionnaire. The reason why water waves were selected as a masker was because the office, wherein the experiment was carried out, was situated near water. The participants may therefore regard water waves as a naturally occurring and less artificial sound in the environment, like the spring water used by Haapakangas *et al.* (2011).

A. Methods

1. Participants

Twenty employees at Norconsult AB [mean age = 34 years, standard deviation (SD) = 9.19] participated in experiment 1. They were recruited by an email invitation sent to all the employees at Norconsult's office in Stockholm, Sweden. Before the experiment proper, participants were asked to report background information such as age, gender, and eventual hearing loss. All participants reported good hearing except for three people who reported a slight hearing impairment. However, the data from these participants did not differ systematically from the rest, and were therefore included in the analyses.

2. Materials

a. Background speech and masking sound. Hagerman sentences (Hagerman, 1982) were used as background speech to simulate a colleague talking in an open plan office. The speech consists of binaural recordings of sentences read by a male actor in Swedish, recorded in an anechoic environment. Each sentence contains four words (names, verbs, etc.) and lasts about 3 s. The entire audio track with the sentences was approximately 5 min with one second pause between each sentence.

Pink noise was created and equalized in the software Audacity to have a 5 dB drop per octave as recommended by Haapakangas *et al.* (2011). The water waves-sound, was designed in the software Cubase Elements 7 and was a mix of two different recordings. The first recording was the sound of ocean waves crushing against land which gave the sound a dynamic. The second recording added a constant and static background sound from ocean wind. The multiple voices sound was based on a recording of nine people talking simultaneously. The recording was multiplied five times to get a more diffuse sound without any distinguishable words or voices. The sounds were saved in WAV format (44.1 kHz, 16 bits). The three masking sounds were calibrated to be played with a sound pressure level of 45 dBA in the two test workstations while the background single voice speech was played at 42 dBA. During the experiment the single voice was masked by the masking sounds, with a total sound pressure level of 47 dBA. Table I summarizes acoustic characteristics of the experimental conditions. Speech Transmission Index (STI) was measured using the indirect method from measured impulse response and derived signal-to-noise ratio (SNR) according to ISO 3382-3:2012 (ISO, 2012) and IEC 60268-16 (IEC, 2011). The octave sound pressure levels for the speech signal were computed by means of the real speech level (IEC, 2011). The speech signal was divided in 18 ms segments of which segments where the A-weighted power exceeding the A-weighted root mean square (rms)

TABLE I. Speech transmission index (STI), STI based on short-time SNRs (sSTI), and mean sound pressure level (SPL) for the experimental conditions in experiments 1 and 2 respectively.

Acoustic measure	Condition			
Experiment 1	Single voice	Single voice + pink noise	Single voice + water waves	Single voice + multiple voices
STI	0.69	0.46	0.38	0.38
sSTI	0.64	0.45	0.39	0.40
SPL (dBA)	42	47	47	47
Experiment 2	1 voice	3 voices	5 voices	7 voices
STI[a]	—	0.47	0.39	0.34
sSTI	—	0.73	0.61	0.49
SPL (dBA)	63.0	67.4	69.6	71.1

[a]The STI values were computed for each voice with the other concurrent voices as noise. The reported values refer to the maximum voice-specific STI value for each condition.

level minus 14 dB were averaged. For the maskers the equivalent octave levels were computed. The effects on speech intelligibility with modulation masking are outside the scope of STI. To consider masker modulation a short-time SNR was computed as the mean of octave SNRs for the 18 ms time segments. Because STI describes intelligibility within 0 and 1 as a linear relation between the apparent SNR within -15 and 15 dB, the 18 ms segments SNR were minimized to 15 dB and maximized to -15 dB, respectively. Acoustic privacy was calculated as "1 − STI" and resulted in the following values: single voice 0.31, pink noise as a masker 0.54, water waves as a masker 0.62 and multiple voices as a masker 0.62.

b. Serial recall task. The serial short-term memory task comprised the immediate recall of sequences of visually presented digits. The digits were presented on a computer screen, in black against a white background with a black frame. Every sequence comprised eight digits randomly selected from the set 1–9 without repetition. Each digit was visible for 500 ms, and the inter-stimulus interval, between the offset of the previous digit until the onset of the next digit, was 300 ms. Half a second after the offset of the last digit in the sequence, the participants were required to reproduce the most recent sequence of digits. They were told to reproduce the digits in the same order as they were presented. Recall was self-paced and undertaken by pressing the corresponding keys on the computer keyboard. When the participants had finished recalling a sequence, they pressed a button to allow for the next sequence to be presented. There were a total of 9 to-be-recalled sequences in each sound condition, summarizing to a total of 45 sequences and there was a brief break between the sound conditions.

c. Questionnaire for subjective ratings of sound. For every sound condition, after the serial recall task was completed, participants were requested to complete a questionnaire with 14 questions/statements (Haapakangas *et al.*, 2011; Haka *et al.*, 2009). The participants used a five-point Likert scale to indicate how much they agreed or disagreed to the given statement. The variable "acoustic satisfaction" was created from 11 statements ("the sound environment was pleasant," "the sound environment was disturbing," "the sound environment was acceptable," "the sound environment was loud," "overall, I was satisfied with the sound environment," "habituation to the sound environment was easy," "surprising changes occurred in the sound environment," "the sound environment often caught my attention," "I could work uninterrupted during the test," and "I could work effectively during the test") and the variable "subjective workload" was created from three statements ("the sound environment impeded my ability to concentrate," "the sound environment impaired my performance," and "the task felt difficult") by calculating the average scores.

3. Design and procedure

A within-participants design was used. There were five sound conditions: silence, single voice, single voice masked by pink noise, single voice masked by water waves, and single voice masked by multiple voices. The participants first undertook the serial recall task in the five sound conditions (with a brief pause between conditions) and, second, responded to the questionnaire for subjective ratings of the sound. Each experimental session lasted about 30 min and the order of the five conditions was counterbalanced by a Latin Square design between the sessions. The experiment took place in an open-plan office measured 10×5 m with a ceiling height of 5 m. The room had a wooden floor, concrete walls and 16 absorbing panels hanging about 1 m from the ceiling covered with Träullit-absorbents, a cement-bonded wood wool material. Ten workstations were located in the room; four of them were used in the experiment. Two of these four places were used for the participants. Two participants were tested at the same time. The single voice speech was played in a Genelec 1029A loudspeaker placed 1.2 m above the floor between the two workstations on opposite side of a 1.4 m high partition screen. The masking sounds were played in two SMS-STR loudspeakers faced upward hanging in the ceiling above the used workstations. The reverberation times in the room were measured to 0.87, 0.65, 0.45, 0.38, 0.40, and 0.38 s in the octave bands 125, 250, 500, 1000, 2000, and 4000 Hz, respectively.

B. Results

1. Serial recall performance

The serial recall task was scored using a strict serial recall criterion: One point was assigned to each item that was recalled in the accurate list position, no other responses received points. As can be seen in Fig. 1, recall performance was worst in the single voice condition and best in the silent and multiple-voice masking conditions. This was confirmed by a repeated measures analysis of variance across the five conditions, $F(4, 76) = 4.71$, $p = 0.002$, $\eta_p^2 = 0.20$. Most importantly, recall was poorer in the "single voice" condition in comparison with the "silent" condition, $t(19) = 3.18$, $p = 0.005$, and in comparison with the "single voice plus multiple voices" condition, $t(19) = 3.21$, $p = 0.005$. There

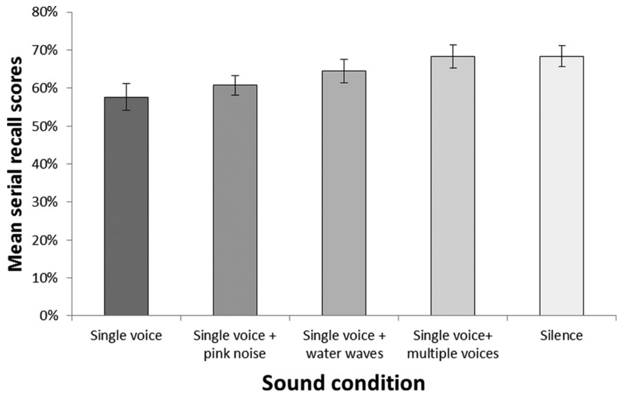

FIG. 1. Mean serial recall scores (percentage correct) across five sound conditions in experiment 1: single voice, single voice plus pink noise, single voice plus water waves, single voice plus multiple voices, and silence. Error bars represent standard error of means.

was no difference between the "silent" and "the single voice plus multiple voices" condition, $t(19) = 0.03$, $p = 0.978$. This indicates that the multiple-voice masker restores error rates to baseline level. Pink noise did not bring back error rates to baseline, as performance was better in the "silent" condition in comparison with the "single voice plus pink noise" condition, $t(19) = 3.09$, $p = 0.006$. Performance in the "single voice plus water waves" condition was, however, not statistically different from performance in the "silent" condition, $t(19) = 1.36$, $p = 0.204$, nor from the "single voice plus multiple voices" condition, $t(19) = 1.26$, $p = 0.222$. Water waves and multiple voices appeared to be best in reducing the detrimental effects of the single voice.

2. Subjective ratings

As can be seen in Fig. 2, acoustic satisfaction was highest in the "silent" condition and lowest in the "single voice" condition, while there was no difference between the three masking conditions. A repeated measures analysis of variance with acoustic satisfaction as the dependent variable revealed a significant effect of condition, $F(4,76) = 17.16$, $p < 0.001$, $\eta_p^2 = 0.47$. Acoustic satisfaction was better in silence compared to any of the other sound conditions, "single voice," $t(19) = 8.82$, $p < 0.001$, "single voice plus pink noise," $t(19) = 5.30$, $p < 0.001$, "single voice plus water waves," $t(19) = 4.43$, $p < 0.001$, and "single voice plus multiple voices," $t(19) = 4.45$, $p < 0.001$. More importantly, acoustic satisfaction was worse in the "single voice" condition compared to "single voice plus water waves" condition, $t(19) = 3.06$, $p = 0.006$, and the "single voice plus multiple voices," $t(19) = 2.19$, $p = 0.041$. No differences were found between the "single voice" condition and the "single voice plus pink noise" condition, $t(19) = 1.47$, $p = 0.159$, or between the "single voice plus multiple voices" condition and the "single voice plus water waves" condition, $t(19) = 0.75$, $p = 0.464$. Subjective workload was also lowest in the silent condition and highest in the "single voice" condition (Fig. 2). A repeated measures analysis of variance with subjective workload as dependent variable found an effect of sound condition, $F(4,76) = 16.23$, $p < 0.001$, $\eta_p^2 = 0.46$. Subjective workload was higher in the "single

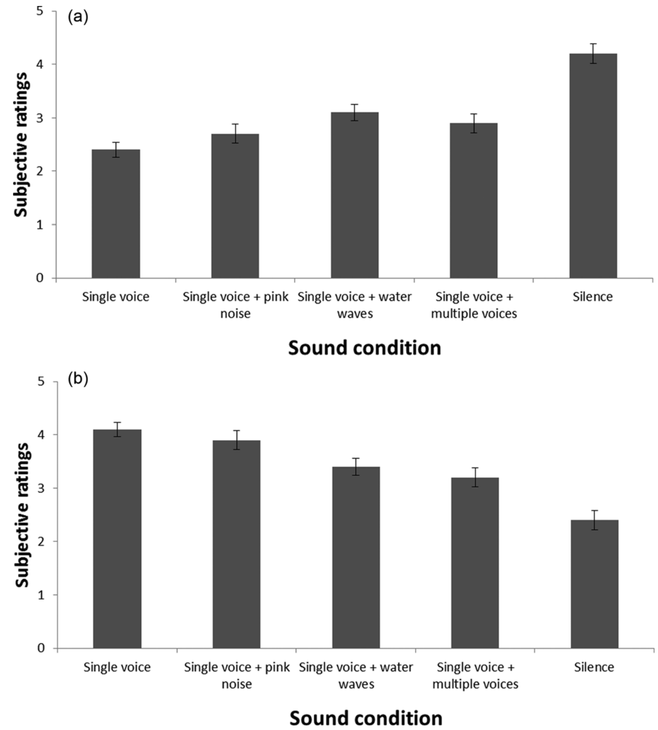

FIG. 2. Mean acoustic satisfaction (panel A) and mean subjective workload (panel B) as estimated on a scale ranging from 1 to 5, in the five sound conditions in experiment 1. Error bars represent standard error of means.

voice" condition as compared with both the "single voice plus water waves" condition, $t(19) = 3.11$, $p = 0.006$, and the "single voice plus multiple voices" condition, $t(19) = 4.56$, $p < 0.001$. Subjective workload was also higher in the "single voice plus pink noise" condition in comparison with the "single voice plus multiple voices" condition, $t(19) = 2.92$, $p = 0.009$. No other comparisons were significant. Overall, then, multiple voices and water waves were perceived as more pleasant maskers than pink noise.

C. Discussion

The results show that serial recall performance was best in silence and in the "single voice plus multiple voices" condition and worst in the "single voice" condition. While water waves also turned out to be a potent masker, in reducing error rates on the serial recall task, multiple voices removed the disruptive impact of irrelevant sound on serial recall performance entirely. Other studies using a comparably low number of voices as the masking sound have merely attenuated the effect of irrelevant sound on serial recall, not removed the effect entirely (Jones and Macken, 1995). As shown here, the effect is removed with a large enough number of voices, possibly because the cues to segmentation (such as abrupt changes in pitch and amplitude) within the auditory stream are eliminated.

In terms of subjective ratings, highest acoustic satisfaction and lowest workload were experienced in silence whereas the lowest acoustic satisfaction and highest workload were experienced in the single voice condition. A comparison of the three masking conditions showed that the acoustic environment was most satisfying when the single voice was masked by water waves or by multiple voices and the lowest workload was experienced in the multiple-voice masking condition. Taken together, multiple voices seem to be the best type of masker, when both performance and subjective ratings are taken into consideration.

III. EXPERIMENT 2

While experiment 1 showed that multiple voices are a particularly potent masker, in terms of protecting performance, experiment 2 explored the characteristics of this type of masker in more detail and within the context of word processed writing instead of the short-term memory task of experiment 1. Four sound conditions were compared in experiment 2: one voice, three voices, five voices, and seven voices. In previous experiments on multiple-voice masking, the sound pressure level has been held constant between conditions (Jones and Macken, 1995). While this approach has its advantage by making it impossible to attribute any effect to differences in sound pressure level, the approach lowers the ecological validity of the experiment as sound pressure level increases when the number of voices in the acoustic environment increases. To obtain a more ecologically valid approach, sound pressure level was higher in conditions with more voices in experiment 2. If sound pressure level is the main factor determining disruption, writing performance should be lowest in the "seven voices" condition and highest in the "one voice" condition. Conversely, if the intelligibility of the background speech is the main factor determining disruption, lowest performance should be found in the "one voice" condition and highest in the "seven voices" condition.

A. Methods

1. Participants

A total of 54 students (mean age = 24.75 years, $SD = 4.34$) at the University of Gävle participated in the study. All participants had completed Swedish compulsory school and high school and all participants had normal or corrected to normal vision. All participants reported normal hearing. The participants received a cinema ticket for their participation.

2. Materials

a. Background speech and multiple-voice maskers. The background speech consisted of audio samples from seven female audiobook narrators and was constructed as follows. Short samples (30–60 s) from monophonic nonfiction audiobooks (MP3 48 kbps CBR 22.1 kHz 32-bit) were converted to WAV format (22.1 kHz 32-bit) and normalized in terms of A-weighted level. The level was computed considering only the active parts of the speech signals (real speech level, IEC 60268-16). The audio samples were, for each narrator, merged into 5-min-long speech signals. Four multiple-voice masker conditions (one voice, three voices, five voices, and seven voices) were composed: "1 voice" condition (i.e., voice 1_1), "3 voice" condition (voice 1_2 + voice 2_1 + voice 3_1), "5 voice" condition (voice 1_3 + voice 2_2 + voice 3_2 + voice 4_1 + voice 5_1), and "7 voice" condition (voice 1_4 + voice 2_3 + voice 3_3 + voice 4_2 + voice 5_2 + voice 6_1 + voice 7_1). The subscript denotes the 1–4 different speech signals created per voice. Because the level was normalized between voices, the sound pressure level increased as a function of voices (i.e., highest sound pressure level in the "7 voice" condition and lowest in the "1 voice" condition). The acoustic characteristics of the conditions are described in Table I. The STI values were computed based on octave SNRs for each voice with the other concurrent voices as noise. Due to the sample rate of 22.1 kHz the 8 kHz octave was implemented as a high-pass filter. As in experiment 1, short-time SNRs were also derived from 18 ms segments. Jorgensen *et al.* (2013) have developed an envelope-power based model for speech intelligibility (SNR_{env}) that correlates to speech reception threshold measurements for modulating maskers, e.g., concurrent speech. The SNR_{env} for the different conditions are presented in Fig. 3. The overall SNR_{env}, rms over the modulation filters 1–64 Hz (second order bandpass filters with octave spacing) and peripheral filters 62.5–6.3 k Hz (gammatone filters with equivalent rectangular bandwidths), were 21.4 dB for three voices, 8.9 dB for five voices, and 6.7 dB for seven voices. The values refer to the maximum voice-specific SNR_{env} value. As it is well documented that writing performance is best in silence compared to noise conditions (e.g., Ransdell *et al.*, 2002; Sörqvist *et al.*, 2012), no silent condition was included in the current experiment to avoid introducing unnecessary error variance.

b. Writing task. The writing task was the same as in Keus van de Poll *et al.* (2014). Participants were asked to write four stories in response to a prompt word that was presented on a computer screen. They were asked to write as much as they could, but to emphasize both speed and accuracy. The prompt words had the names of different nature scenes (for example, mountains and field). The prompt words were presented in the same sequential order to all participants. The onset and offset of the prompt words and the sounds were synchronized. The time limit was set to 5 min for each story. After 5 min, a voice instructed the participant to stop writing, and move on to the next condition. The computer software program SCRIPTLOG (version 1.8.19, January 2005) was used to obtain data. This program registers all keyboard activity and is developed for real-time analyses of the writing process. The dependent measure used in the analyses was writing fluency (i.e., the total number of characters in the final edited text plus the total number of deleted characters).

3. Design and procedure

A within-subjects design was used. Each participant sat in front of a stationary computer in a sound isolated room. They wore headphones during the whole experiment. The writing task was introduced by a practice phase of 30 s so the participants would get acquainted with the task and the procedure. The prompt-word "city" was presented in the practice phase. The practice phase was followed by the four sound (number of voices) conditions. The order of the four

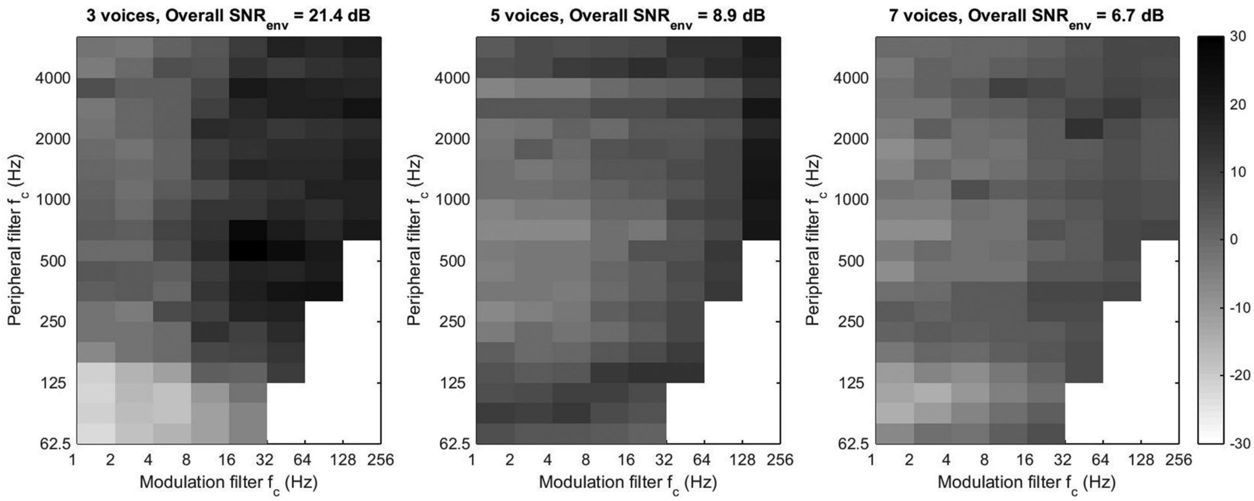

FIG. 3. SNR$_{env}$ in dB for the three conditions three voices (left), five voices (middle), and seven voices (right). The values refer to the maximum voice-specific SNR$_{env}$ values and are in gray scale as indicated by the scale on the right.

conditions was counterbalanced across participants. The participants were told to ignore any sounds presented over headphones. The experiment took about 30 min to complete.

B. Results

Writing fluency increased as a function of number of voices in the background, with the highest performance in the condition with most voices and lowest performance in the condition with only one voice (Fig. 4). This was statistically confirmed by a repeated measures analysis of variance with sound condition as independent variable, $F(3, 156) = 5.13$, $p = 0.002$, $\eta^2_p = 0.09$. Of key interest, a contrast analysis showed that the linear trend across conditions was significant, $F(1, 52) = 15.48$, $p < 0.001$, $\eta^2_p = 0.23$. Follow up *t*-tests showed that writing fluency was significantly lower in the condition with one voice than in the condition with three voices, $t(52) = 2.17$, $p = 0.034$. There was also a significant difference between one and five voices, $t(52) = 2.37$, $p = 0.022$, between one and seven voices, $t(52) = 4.35$, $p < 0.001$, and between three and seven voices, $t(52) = 1.79$, $p = 0.04$ (one-tailed). All other comparisons were non-significant. It should be noted that these multiple comparisons run the risk of type I errors. If a Bonferroni correction is applied to check the results against inflation of alpha error, the only two conditions that differ significantly are the one-voice condition and the seven-voice condition. However, linear function between the number of voices and writing fluency is not concerned with this inflation and Bonferroni corrections enhance the risk for type II errors. Taking these considerations into account, the Bonferroni correction does not change the main implications of the results.

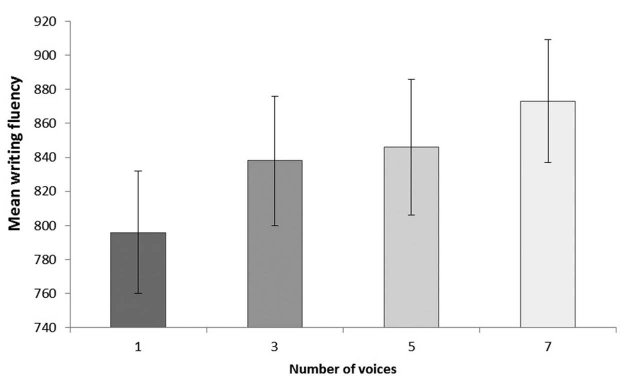

FIG. 4. Mean writing fluency (characters in the final edited text plus deleted characters during the writing process) scores in experiment 2, across four sound conditions with a various number of voices talking simultaneously in the background (one voice, three voices, five voices, and seven voices). Error bars represent standard error of means.

C. Discussion

Writing fluency increased when more voices were added in the background. The concomitant reduction in the disruption produced by babble harmonises with the finding that a reduction in STI (or an increase in "privacy"), enhances writing performance (Keus van de Poll *et al.*, 2014). This effect is unlikely to be attributable to the mechanism responsible for the changing-state effect, as writing is not sensitive to mere acoustic change but sensitive to disruption from the meaning of background speech (Sörqvist *et al.*, 2012). Whereas masking reduces the magnitude of the changing-state effect, in the context of serial recall (experiment 1), by smoothing the distinct elements in the speech signal into a more continuous noise, the effects of masking in experiment 2 is probably due to reduced speech intelligibility. A systematic reduction of the intelligibility of the speech signal, toward a non-intelligible babble, reduces the disruptive impact of the semantic properties of irrelevant sound on writing performance.

IV. GENERAL DISCUSSION

Experiment 1 found that serial recall performance is impaired by a single voice, but when this single voice is masked by multiple voices, error rates return to baseline. Multiple-voice masking was also, together with water waves, evaluated as the most pleasant masker, in terms of workload and acoustic satisfaction. Based on these findings, experiment 2 was designed to study the effects of multiple-voice masking in more detail, and to do so within the context of an arguably more applied type of task. Experiment 2 found that writing performance is lower with one voice in the background in comparison with more voices, even though sound pressure level was highest in the seven voice condition and lowest in the one voice condition.

A. Why masking protects performance

One means by which the potential distraction from irrelevant sound can be overcome is to add a masking sound to the speech signal. Adding a masking sound can achieve a double duty: Not only does it reduce the intelligibility of the irrelevant sound, it also reduces its temporal spectral characteristics, thereby reducing the cues required to perceptually segment the sound. In experiment 1, we demonstrated that performance was best in silence and in "the single voice plus multiple voices" condition and was poorest in the single voice condition. Here then, "multiple voices" was the most effective masker and pink noise the least effective masker. This is consistent with the view that the addition of many voices smoothes the peaks and troughs of the acoustic signal such that the abrupt pitch and loudness changes become more temporally constant. The cues to segmentation, and therefore the changing-state of the sound, are effectively reduced, thereby preventing the automatic encoding on the order of changing-state and attenuating the irrelevant sound effect.

In experiment 2, a writing task was used as the evaluative instrument instead of serial recall and performance was found to be negatively related to number of voices in the background. Disruption to writing was not a result of sound intensity. If sound intensity was a property of background speech that produces disruption, then performance should have been lowest in the "7 voices" condition. However, performance was highest in the "7 voices" condition, and lowest in the "1 voice" condition. This is in line with a raft of previous findings that demonstrate that the disruption produced by irrelevant sound of serial recall is not dependent on sound intensity (Colle, 1980; Jones *et al.*, 1990).

Another possibility is that the effects seen in experiment 2 are attributable to reduced changing-state characteristics of the background sound, just as in experiment 1. In this case, as the number of voices increased, the changing-state features of the background sound were reduced, and performance increased. However, this explanation is not compatible with previous research on the effects of sound on word processed writing. This task is known to be peculiarly sensitive to disruption via the semantic properties of irrelevant sound while being invulnerable to its acoustic properties (Keus van de poll *et al.*, 2014; Sörqvist *et al.*, 2012). Taken together, the masking effects found in experiment 2 are best explained as a reduction of the background sound's intelligibility: As the number of voices in the background increased, intelligibility decreased, thereby impairing the automatic processing of meaning and reducing the interference of such processing with the deliberate semantic processes underpinning the writing task. In experiment 2, both semantic and acoustic properties of the sound signal were manipulated, and hence confounded, in an attempt to study the effects of multiple-voice masking from an applied viewpoint at the cost of theoretical ambitions.

B. The effectiveness of different maskers

A multitude of voices was sufficient to completely remove the disruptive effect of irrelevant sound on performance (experiment 1). One take on this effect is that voices are optimal as a masker. Failures to eliminate the effect of irrelevant sound entirely may be because additional voices did not exceed 8 in number (Ebissou *et al.*, 2013; Hellbrück and Kilcher, 1993; Jones and Macken, 1995). However, caution should be exerted here since babble produced by 100 speakers has been shown to produce significant disruption relative to quiet (Hellbrück and Kilcher, 1993) suggesting that it may be important to take into account the acoustic methods used to create the speech samples.

Water waves were also identified as a potent masker (experiment 1), ameliorating the disruptive effect of background speech better than broadband noise. It is possible that this may be for acoustic reasons: by being a better masker of the acoustic features of the single voice embedded in the background sound. However, water waves may also exert its advantage by being a more "natural" sound as opposed to broadband noise. This is consistent with the results by Haapakangas *et al.* (2011) who found that spring water is a sufficient masker for the effects of background speech on serial recall. This may also be reflected in the acceptance rates for the water sound.

Speech Transmission Index (STI) was used to describe the conditions in the two experiments as in accordance to standardized measurements of acoustic parameters for open plan office (ISO, 2012). To include the effects of modulation masking short-time STI values were also computed based on average 18 ms signal-to-noise ratios (SNRs). The measure is not validated; here it is used to illustrate the effects of masker modulation. A short-time speech-based STI has been described by Payton and Shrestha (2013). An extension of the speech intelligibility index (SII) based on short time segments (12 ms) has also been proposed (Rheberger and Versfeld, 2005). In experiment 1 the maskers were concluded stationary; the short-time STI produced the same result (see Table I). In experiment 2, the short-time STI measure showed significantly higher speech intelligibility for the three voice condition. This is further corroborated by the envelope-power based SNR (Jorgensen *et al.*, 2013) reported in Fig. 3.

C. Performance in comparison with subjective measures

The performance and subjective measures mismatched in experiment 1. There was no difference in short-term memory performance between the "silent" condition and the "single voice plus multiple voices" condition. However, this was not the case for subjective acoustical satisfaction and workload: Silence was evaluated as a more pleasant acoustic environment than the "single voice plus multiple voices" condition. So even if performance measures indicate that working in open offices is good as long as there are a lot of voices in the background, instead of just one, the subjective ratings indicate that individuals prefer quiet environments for cognitive work (see also Haapakangas *et al.*, 2011; Jahncke *et al.*, 2013; Schlittmeier and Hellbrück, 2009). Nonetheless, multiple voices were evaluated as a relatively pleasant masker, which is consistent with the performance data and suggests that multiple voices are the best masker in terms of protecting performance while compromising acoustic satisfaction to a limited extent.

Although in the present experiment 2 exclusively performance date was collected, the results of extant studies suggest, that subjective measures already reveal subtle lowering in STI (Haapakangas *et al.*, 2011; Haka *et al.*, 2009) before these start to be also reflected in performance measures. In fact, tasks become rated as more difficult even at STIs of 0.35 (Haka *et al.*, 2009), which are assumed to be acceptable within the open plan office—at least with respect to performance (ISO, 2012).

D. Practical implications of masking noise within the office

Intelligible background speech (high STI values, low privacy values) is the outstanding noise problem in open-plan offices, in which employees not only communicate with each other, but also need to concentrate and perform cognitive activities preferably in silence. Our study speaks in favor of reducing background speech intelligibility to protect performance, by adding multiple voices to the acoustic environment, or when the type of activity in the office (and its size) permits, by increasing (rather than decreasing) the number of co-workers who contribute to the sound in the acoustic environment. A large number of voices protects performance and increases privacy. Such a potential beneficial effect of several background voices compared to a single speaker is mentioned in the recently published ISO Standard 3382-3. However, ISO 3382-3 refers to the possibility that simultaneously talking colleagues partially mask each other; a situation which might occur only by chance (e.g., a sufficiently high number of colleagues talking from the approximate spatial location). The cost, however, is impaired communication. Workspace satisfaction depends on a communication/privacy tradeoff in open-plan offices (Kim and de Dear, 2013) and a simple solution that may satisfy both communication and privacy demands is difficult to reach in practice. Attempts at this have been made (e.g., the Babble Voice Privacy System by Herman Miller) but we are unaware of scientific studies that back up their effectiveness. We consider it a highly appreciated engineering task to develop technical masking solutions that satisfy these requirements.

ACKNOWLEDGMENT

Experiment 1 was part of J.C.'s graduate thesis. We would like to thank Mattias Holmgren and Andreas Haga for assistance with the data collection.

Beaman, C. P., and Holt, N. J. (**2007**). "Reverberant auditory Environments: The effects of multiple echoes on distraction by 'irrelevant' speech," Appl. Cogn. Psychol. **21**, 1077–1090.

Bregman, A. S., and Rudnicky, A. I. (**1975**). "Auditory segregation: Stream or streams?," J. Exp. Psychol.: Human Percept. Perform. **1**, 263–267.

Colle, H. A. (**1980**). "Auditory encoding in visual short-term recall: Effects of noise intensity and spatial location," J. Verbal Learn. Verbal Behav. **19**, 722–735.

Ebissou, A., Chevret, P., and Parizet, E. (**2013**). "Work performance and mental workload in multiple talker environments," in *International Congress on Acoustics 2013*, December 2012, Montréal, Canada, No. 4pNSa8.

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

Haapakangas, A., Kankkunen, E., Hongisto, V., Virjonen, P., Oliva, D., and Keskinen, E. (**2011**). "Effects of five speech masking sounds on performance and acoustic satisfaction. Implications for open-plan offices," Acta Acust. Acust. **97**, 641–655.

Hagerman, B. (**1982**). "Sentences for testing speech intelligibility in noise," Scand. J. Audiol. **11**, 79–87.

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

Halin, N., Marsh, J. E., Haga, A., Holmgren, M., and Sörqvist, P. (**2013**). "Effects of speech on proofreading: Can task-engagement manipulations shield against distraction?," J. Exp. Psychol.: Appl. **20**, 69–80.

Halin, N., Marsh, J. E., Hellman, A., Hellström, I., and Sörqvist, P. (**2014**). "A shield against distraction," J. Appl. Res. Mem. Cognit. **3**, 31–36.

Hellbrück, J., and Kilcher, H. (**1993**). "Effects on mental tasks induced by noise recorded and presented via an artificial head system," in *Noise and Man' 93*, edited by M. Vallet (Institut National de Recherche sur les Transports, Leur Securite, Arcueil, France), pp. 315–322.

IEC (**2011**). IEC 60268-16, *Sound System Equipment—Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index*, 4th ed. (IEC, Geneva, Swizerland).

ISO (**2012**). ISO 3382-3:2012, *Acoustics—Measurement of Room Acoustic Parameters—Part 3: Open Plan Offices* (International Organization for Standardization, Geneva, Switzerland).

Jahncke, H., Hongisto, V., and Virjonen, P. (**2013**). "Cognitive performance during irrelevant speech: Effects of speech intelligibility and office-task characteristics," Appl. Acoust. **74**(3), 307–316.

Jahncke, H., Hygge, S., Halin, N., Green, A-M., and Dimberg, K. (**2011**). "Open-plan office noise: Cognitive performance and restoration," J. Env. Psychol. **31**, 373–382.

Jiang, B., Liebl, A., Leistner, P., and Yang, J. (**2011**). "Sound masking performance of time-reversed masker processed from the target speech," Acta Acust. Acust. **98**, 135–141.

Jones, D. M., and Macken, W. J. (**1993**). "Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory," J. Exp. Psychol.: Learn. Mem. Cognit. **19**, 369–381.

Jones, D. M., and Macken, W. J. (**1995**). "Auditory babble and cognitive efficiency: Role of number of voices and their location," J. Exp. Psychol.: Appl. **1**, 216–226.

Jones, D. M., Marsh, J. E., and Hughes, R. W. (**2012**). "Retrieval from Memory: Vulnerable or Inviolable?," J. Exp. Psychol.: Learn. Mem. Cognit. **38**, 905–922.

Jones, D. M., Miles, C., and Page, J. (**1990**). "Disruption of proofreading by irrelevant speech: Effects of attention, arousal or memory?," Appl. Cognit. Psychol. **4**, 89–108.

Jones, D. M., and Tremblay, S. (**2000**). "Interference in memory by process or content? A reply to Neath (2000)," Psychon. Bull. Rev. **7**, 550–558.

Jorgensen, S., Ewert, S. D., and Dau, T. (**2013**). "A multi-resolution envelope-power based model for speech intelligibility," J. Acoust. Soc. Am. **134**, 436–446.

Keus van de Poll, M., Ljung, R., Odelius, J., and Sörqvist, P. (**2014**). "Disruption of writing by background speech: The role of speech transmission index," Appl. Acoust. **81**, 15–18.

Kim, J., and de Dear, R. (**2013**). "Workspace satisfaction: The privacy-communication trade-off in open-plan offices," J. Environ. Psychol. **36**, 18–26.

Kittel, M., Wenzke, E., Drotleff, H., and Liebl, A. (**2013**). "Auditory babble as a masker of disruptive speech," in *Proceedings of Internoise*, September 15–18, Innsbruck, Austria.

Loewen, L. J., and Suedfeld, P. (**1992**). "Cognitive and arousal effects of masking office noise," Environ. Behav. **24**, 381–395.

Marsh, J. E., Hughes, R. W., and Jones D. M. (**2008**). "Auditory distraction in semantic memory: A process-based approach," J. Mem. Lang. **58**, 682–700.

Marsh, J. E., Hughes, R. W., and Jones, D. M. (**2009**). "Interference by process, not content, determines semantic auditory distraction," Cognition **110**, 23–38.

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

Martin, R. C., Wogalter, M. S., and Forlano, J. G. (**1988**). "Reading comprehension in the presence of unattended speech and music," J. Mem. Lang. **27**, 382–398.

Park, M., Kohlrausch, A., and van Leest, A. (**2013**). "Irrelevant speech effect under stationary and adaptive masking conditions," J. Acoust. Soc. Am. **134**, 1970–1981.

Payton, K. L., and Shrestha, M. (**2013**). "Comparison of a short-time speech-based intelligibility metric to the speech transmission index and intelligibility data," J. Acoust. Soc. Am. **134**, 3818–3827.

Perham, N., Banbury, S., and Jones, D. M. (**2005**). "Auditory distraction impairs analytical reasoning performance," Aust. J. Psychol. **57**, 242–242.

Perham, N., Banbury, S., and Jones, D. M. (**2007**). "Do realistic reverberation levels reduce auditory distraction?," Appl. Cogn. Psychol. **21**, 839–847.

Perham, N., and Currie, H. (**2014**). "Does listening to preferred music improve reading comprehension performance?," Appl. Cogn. Psychol. **28**, 279–284.

Perham, N., and Sykora, M. (**2012**). "Disliked music can be better for performance than liked music," Appl. Cogn. Psychol. **26**, 550–555.

Ransdell, S., Levy, C. M., and Kellogg, R. T. (**2002**). "The structure of writing processes as revealed by secondary task demands," L1-Educat. Stud. Lang. Lit. **2**, 141–163.

Rheberger, K. S., and Versfeld, N. J. (**2005**). "A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," J. Acoust. Soc. Am. **117**, 2181–2192.

Schlittmeier, S. J., and Hellbrück, J. (**2009**). "Background music as noise abatement in open-plan offices: A laboratory study on performance effects and subjective preferences," Appl. Cogn. Psychol. **23**, 684–697.

Schlittmeier, S. J., Hellbrück, J., and Klatte, M. (**2008**). "Does irrelevant music cause an irrelevant sound effect for auditory items?," Eur. J. Cogn. Psychol. **20**, 252–271.

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

Sörqvist, P., Halin, N., and Hygge, S. (**2010**). "Individual differences in susceptibility to the effects of speech on reading comprehension," Appl. Cogn. Psychol. **24**, 67–76.

Sörqvist, P., Marsh, J. E., and Jahncke, H. (**2010**). "Hemispheric asymmetries in auditory distraction," Brain Cognit. **74**, 79–87.

Sörqvist, P., Nöstl, A., and Halin, N. (**2012**). "Disruption of writing processes by the semanticity of background speech," Scand. J. Psychol. **53**, 97–102.

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