International Journal of Psychophysiology 200 (2024) 112352

Contents lists available at ScienceDirect

International Journal of Psychophysiology

journal homepage: www.elsevier.com/locate/ijpsycho

The effects of irrelevant speech on physiological stress, cognitive performance, and subjective experience – Focus on heart rate variability

Jenni Radun[a,*], Henna Maula[a], Iida-Kaisa Tervahartiala[a], Ville Rajala[a], Sabine Schlittmeier[b], Valtteri Hongisto[a]

[a] Psychophysics Laboratory, Turku University of Applied Sciences, Turku, Finland
[b] Work and Engineering Psychology, RWTH Aachen University, Aachen, Germany

ARTICLE INFO

Keywords:
Irrelevant sound effect
Stress
Heart rate variability
Performance
Speech

ABSTRACT

Irrelevant speech impairs cognitive performance, especially in tasks requiring verbal short-term memory. Working on these tasks during irrelevant speech can also cause a physiological stress reaction. The aim of this study was to examine heart rate variability (HRV) as a non-invasive and easy-to-use stress measure in an irrelevant speech paradigm. Thirty participants performed cognitive tasks (n-back and serial recall) during two sound conditions: irrelevant speech (50 dB) and quiet (33 dB steady-state noise). The influence of conditions as well as presentation orders of conditions were examined on performance, subjective experience, and physiological stress. Working during irrelevant speech compared to working during quiet reduced performance, namely accuracy, in the serial recall task. It was more annoying, heightened the perceived workload, and lowered acoustic satisfaction. It was related to higher physiological stress by causing faster heart rate and changes in HRV frequency-domain analysis (LF, HF and LF/HF). The order of conditions showed some additional effects. When speech was the first condition, 3-back performance was less accurate, and serial recall response times were longer, heart rate was faster, and successive heart beats had less variability (lower RMSSD) during speech than during quiet. When quiet was the first condition, heart rate was faster and reaction times in 3-back were slower during quiet than during speech. The negative effect of irrelevant speech was clear in experience, performance, and physiological stress. The study shows that HRV can be used as a physiological stress measure in irrelevant speech studies.

1. Introduction

It is well known that cognitive performance in certain tasks is significantly lower during intelligible but task-irrelevant speech than during silence (Hongisto, 2005; Schlittmeier et al., 2012; Szalma and Hancock, 2011). Task-irrelevant speech means speech that is not related to the task at hand, but is present in the acoustic background. Its presence disturbs performance in visually presented verbal short-term memory tasks (Ellermeier and Zimmer, 1997; Jones and Macken, 1993; Salamé and Baddeley, 1987). Speech impairs performance more than continuous noise at the same sound level (Ellermeier and Zimmer, 1997). Especially the performance decrement in verbal serial recall tasks relates to speech intelligibility (Muhammad et al., 2019; Schlittmeier et al., 2008), and its objective estimate, the Speech Transmission Index (STI) (Haapakangas et al., 2020; Hongisto, 2005). The performance decline in these tasks exceeds 15 % with STIs over 0.6 (Haapakangas et al., 2020). An STI above 0.6 means that the speech signal's semantics are completely understandable. Other tasks influenced by irrelevant speech are, for example, the n-back task, where the accuracy in 3-back task was lower during speech and noise than during quiet (Radun et al., 2021) and reaction time was raised when speech was clearer, i.e., STI was higher (Haapakangas et al., 2014).

This detrimental effect related to irrelevant speech is more generally described as the irrelevant sound effect, where the presence of a task-irrelevant sound sequence disrupts short-term serial recall even when the participants are told to ignore the sound (Ellermeier and Zimmer, 1997; Hughes et al., 2007). This effect is related to temporal-spectral variations in auditory streams, as steady or repetitive tones did not cause the effect (Jones and Macken, 1993). First, the effect was explained by the changing-state hypothesis, which states that both serial task performance and sound sequence processing relies on seriation processes being the crucial factor (Jones and Macken, 1993; Macken

* Corresponding author.
E-mail address: jenni.radun@turkuamk.fi (J. Radun).

https://doi.org/10.1016/j.ijpsycho.2024.112352
Received 22 September 2023; Received in revised form 30 January 2024; Accepted 16 April 2024
Available online 18 April 2024

et al., 1999). Later, the duplex-mechanism account of auditory distraction connects this effect on performance with two distinct mechanisms: interference-by-process and attention capture (Hughes et al., 2007). Interference-by-process means that the disturbance of performance is related to the extent to which the cognitive processing of both irrelevant sound and cognitive task shares same cognitive processes (Hughes et al., 2007). Attention capture describes that auditory events differing from the recent auditory past capture attention and disrupt performance (Hughes et al., 2007). High working memory capacity has been suggested to attenuate the detrimental effects based on attention capture, but not the disturbance effects that originate from interference-by-process (Hughes et al., 2013; Sörqvist, 2010). However, also studies finding no relation between working memory capacity and performance exist (e.g., Körner et al., 2017).

Working during irrelevant speech can also cause stress. Several studies have reported elevated stress hormone concentrations while working during task-irrelevant speech (Evans and Johnson, 2000; Radun et al., 2021; Tafalla and Evans, 1997). However, measuring stress hormone levels is laborious and places several demands on the experimental procedure. Easier and less invasive means of detecting physiological signs of stress exist, but they are still less reported and explored. Examples of this kind of non-invasive physiological stress measures in the context of the irrelevant sound paradigm are pupil size (Marois et al., 2019) and heart rate variability (HRV) (Radun et al., 2021). In this study, HRV will be examined in more detail in an irrelevant sound paradigm.

HRV is suitable for short-term measurements, and it can be calculated from the electrocardiogram (ECG). The measurement does not place as high requirements on the experimental design as hormone level measurements do, since a standard chest belt can collect all the data. HRV reflects physiological stress by examining the functioning of the sympathetic and parasympathetic branches of the autonomic nervous system. The imbalance between these branches, i.e., autonomic imbalance, is usually characterized by a hyperactive sympathetic system and a hypoactive parasympathetic system (Thayer et al., 2010). This imbalance is further related to many factors, such as hypertension, obesity, family history, and work stress, which are also risk factors for cardiovascular disease (Thayer et al., 2010).

For short-term measurements using HRV, frequency-domain analysis provides information on autonomic balance (Kim et al., 2018). The frequency domain analysis divides HRV into spectral components (high-frequency band (HF) 0.15–0.4 Hz, low-frequency band (LF) 0.04–0.15 Hz, and very low frequency band (VLF) <0.04 Hz) (Kim et al., 2018). Stress is most frequently associated with low parasympathetic activity, characterized by a decrease in HF and an increase in LF (Kim et al., 2018). Therefore, a frequently used stress indicator obtained by HRV measurement is the LF/HF ratio. High values denote low parasympathetic and high sympathetic activity, i.e., stress.

Even the mere exposure to sounds (without performing tasks) can cause stress, as indicated by these HRV markers. Lee et al. (2010) examined the effect of sound pressure level (SPL) on HRV by exposing sixteen participants to white noise of 50, 60, 70, and 80 dB L_{Aeq} (A-weighted equivalent SPL) for 5 min. White noise with an SPL of 38 dB L_{Aeq} acted as a reference. The noise exposure increased LF and the LF/HF ratio of HRV, indicating higher stress with higher SPL. However, it influenced neither heart rate, mean heart pressure, nor HF. The LF/HF ratio is additionally positively correlated with the SPL. Exposure to lower SPLs has also been examined. In a between-subjects study, forty participants were divided into four groups, which were exposed to sound conditions speech, traffic, traffic and speech (all 45 dB L_{Aeq}), and quiet (35 dB L_{Aeq}) (Sim et al., 2015). The exposure to speech reduced the LF value compared to the quiet condition, indicating that speech created less stress than the other sound conditions. These participants were merely exposed to sounds without performing tasks; therefore, hearing speech might be more comfortable than exposure to other sound types in the experiment. The sound exposure, therefore, can influence HRV, but the effects of task-irrelevant sounds, especially speech, might differ if one performs cognitive tasks during the sound exposure.

The effects of sound exposure during cognitive tasks depend on the type of sounds and tasks. HRV was examined in ten participants performing cognitive tasks related to office drawing and working memory order judgement tasks during office noise containing speech at SPLs of 65 dB L_{Aeq} and 41 dB L_{Aeq} (Kristiansen et al., 2009). The sound conditions did not affect physiological stress (HRV LF, HF, LF/HF, and total power) nor performance. Only the mental load imposed by the tasks was seen as an increase in LF and LF/HF ratio compared to the rest condition when no sounds or tasks were presented. However, even the quietest sound condition involved intelligible speech, which may explain the lack of effect of the sound condition. It is well known that the influence of speech on performance is more related to speech intelligibility than the sound level, when speech is audible (Haapakangas et al., 2020; Hongisto, 2005; Schlittmeier et al., 2008).

The effect of speech on cognitive performance is larger than the effect of non-speech noise (Szalma and Hancock, 2011). Especially simple verbal short-term memory tasks, such as the serial recall task, are sensitive to speech effects (Haapakangas et al., 2020). Though the verbal serial recall task is a simple task of remembering the order of unrelated verbal items (e.g. words, syllables, digits), it requires verbal short-term memory, which is necessary in most tasks related to studying and office work. Radun et al. (2021) compared the physiological effects of speech (65 dB L_{Aeq}), steady-state wideband sound at the same SPL (65 dB L_{Aeq}), and quiet steady-state sound (35 dB L_{Aeq}) while performing short-term memory tasks using a between-subject design (three groups). They found that over time, the LF/HF ratio rose more in the speech group than in the other two groups, suggesting higher stress over time when working during speech than during steady state sounds. As the individual differences in HRV are substantial, and a within-subject design is more sensitive than a between-subjects design, a detailed analysis of the effects of speech on HRV utilizing a within-subject design is necessary to estimate the effects of intelligible speech on a working person.

Our first aim was to make a controlled study on the within-subject effects of speech while working on tasks sensitive to the effects of speech, with a focus on the stress measure HRV. Two sound conditions were compared: task-irrelevant speech and quiet. The speech condition featured separate sentences irrelevant to the tasks performed (50 dB). The level corresponds to a typical level of speech measured in open-plan offices (Yadav et al., 2021). In offices, the task-irrelevant speech is often naturally occurring, intelligible speech consisting perhaps of separate sentences or partly heard dialogues. Therefore, our speech was selected to consist of sentences with predictable timing and content spoken with the same male voice. In the terms of the duplex-mechanism account of auditory distraction, this kind of material should rather interfere with specific task-related processes than capture attention (Hughes et al., 2007). However, as we used isolated sentences, which were grammatically correct, highly intelligible and semantically meaningful, the semantic unpredictability of certain sentences may sometimes capture attention (cp. Marsh et al., 2018). The quiet condition was soft and steady office ventilation noise (33 dB). The effects measured were experience, performance, and physiological stress measured with HRV. We wanted to test whether working during speech shows differences in HRV that can relate to increased stress when compared with working during quiet.

In our experimental design, we aimed to ensure that our observed effect was attributed to the sound condition rather than other factors. Some studies on the irrelevant sound effect have found that performance in first trials or sessions might be poorer than in later trials or sessions (Ellermeier and Zimmer, 1997; Farley et al., 2007; Hellbrück et al., 1996; Röer et al., 2014). Despite a rehearsal session before the actual experiment, some studies have noted a practice effect where participants performed worse during the first one or two experimental sessions in all sound conditions (Ellermeier and Zimmer, 1997; Hellbrück et al., 1996). Despite this practice effect, the effect of the irrelevant sound remained stable within the sessions (Ellermeier and Zimmer, 1997; Hellbrück et al., 1996). Others have identified a habituation effect to complex sound distractors, indicating that the performance is poorer during the first trials of a serial recall task, and this effect is not apparent in a quiet condition (Röer et al., 2014). This habituation effect was related to complex auditory distractors such as speech and to the auditory capture that habituation can attenuate (Röer et al., 2014).

Therefore, as we opted for a within-subject design, we also aimed to examine the effect of presentation order, namely the order in which the sound conditions were presented to the participants. A previous study examining serial recall performance in quiet and speech conditions demonstrated that the effect of speech was found in both between-subject and within-subject designs, but in the latter, an interaction between presentation order and sound condition emerged as well (Farley et al., 2007). Although no main effect of presentation order was found, the recall accuracy during task-irrelevant speech was better when the speech was presented after the quiet condition compared to the opposite order (Farley et al., 2007). Prior studies have not investigated whether this interaction effect would also be visible in physiological stress measured with HRV.

2. Methods

2.1. Design

The study was conducted as a psychological laboratory experiment with a 2 × 2 design (2 *conditions* and 2 *presentation orders* of conditions). The *conditions* were speech and quiet. The *condition* served as a within-subject variable, while the *presentation order* was a between-subjects variable. To ensure that the compared *presentation order* groups were comparable in important aspects that could potentially influence performance, a few background factors were controlled for. Since some studies have suggested high working memory capacity to mitigate attention capture during irrelevant speech (Hughes et al., 2013; Sörqvist, 2010), and we cannot rule out the possible attention capture of the speech, the working memory capacity of participants in different presentation order groups was assessed. Additionally, noise might impact the performance of noise-sensitive individuals more than that of non-noise-sensitive individuals (Belojević et al., 1992; Ellermeier and Zimmer, 1997). Therefore, the noise sensitivity of the groups was also examined.

2.2. Participants

Participants were primarily recruited through student email lists. The inclusion criteria were native-level Finnish speaker, having normal hearing, and falling within the age range of 18 to 45 years. A total of 30 participants took part in the study, with an average age of 24 years (min. 19 years and max. 42 years). Among the participants, ten were male.

The number of participants was determined before the experiment using power calculations. The main anticipated effect was the repeated-measure effect of the *condition,* with no expected between-subjects effect. The intraclass correlation (ICC2) can be used to raise the effect size. For instance, if ICC2 = 0.8, then a moderate effect size of D = 0.4 could be raised to a large effect size of D = 0.6 (Brysbaert, 2019). Our primary outcome variables were performance in a serial recall task and the HRV LF/HF. We examined the ICC2 in the quiet condition from one of our previous experiments ($N = 98$). For the 10 trials of the serial recall task, ICC2 = 0.791, and for the HRV LF/HF ratio for three task, ICC2 = 0.87. Therefore, with an effect size of D = 0.6, a power of 0.8, and $p < 0.05$, the required minimum number of participants was 24 using 2 × 2 split-plot design, with the main effect being the repeated measure (Brysbaert, 2019). Due to the possibility of outliers and potential HRV recording failures for some participants, we decided to recruit 30 participants.

2.3. Ethical aspects

The Ethical Committee of the Turku University of Applied Sciences approved the study on April 28, 2020 [1/2020]. All participants provided voluntary, informed consent before participating. Participants were compensated for their time and effort with a gift voucher worth 20 Euro.

2.4. Laboratory

The experiment was conducted in the Psychophysics Laboratory of Turku University of Applied Sciences, Turku, Finland, within October–December 2020. The layout drawing of the experimental room is shown in Fig. S1 in the Supplementary material. The background noise level of the room, while air-conditioning was operational, was under 17 dB L_{Aeq} falling below the human hearing threshold within the frequency range 20–10,000 Hz. Air quality was maintained at the highest standards throughout the experiment. The temperature of the room during the experimental sessions ranged between 22 °C and 24 °C. The fresh air inlet rate was 30 l/s ensuring a low concentration of CO_2. The illumination level on the table in front of the participants was approximately 500 lx, which was suitable for computer work and did not cause any glare.

2.5. Experimental conditions

The quiet condition consisted of steady-state wideband background noise. The measured A-weighted equivalent SPL of the sound at the listener's head position was 33 dB L_{Aeq}. The measured spectrum of the sound is shown in Fig. 1.

The quiet condition resembled a typical background noise generated by ventilation systems in offices. We opted not to test absolute silence (e.g., 15 dB L_{Aeq}), as it is unrealistic in real working environments. We employed a commercial sound masking system to create the quiet condition. The system comprised a control unit (Cambridge Sound Management Qt 100, Biamp Systems LLC, Beaverton, Oregon, USA) and wires to four loudspeakers (Qt emitter, diameter 83 mm) which were embedded in the suspended ceiling (40 mm mineral wool) at a height of 2750 mm. The horizontal distribution of the sound pressure level at the height of the head was very even, and the masking sound sources could

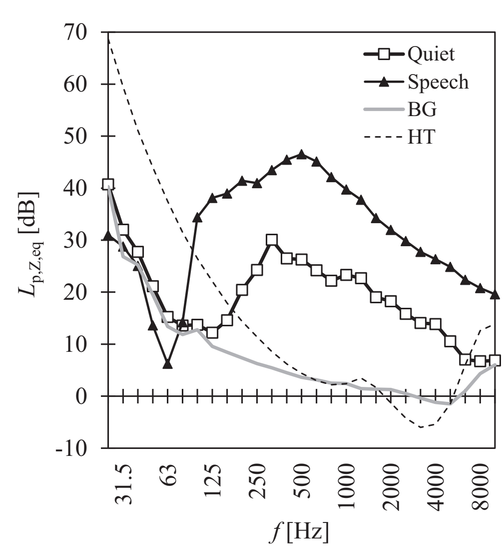

Fig. 1. The unweighted one-third-octave equivalent sound pressure level, $L_{p,Z,eq}$, as a function of the frequency of sound, f, for the *conditions* Quiet and Speech within 20–10,000 Hz. In addition, the background noise level (BG) of the experimental room and the average hearing threshold level (HT) of a normal-hearing adult according to ISO 226 standard (ISO, 2003a) are indicated.

not be located by hearing. The loudspeakers were white, blending with the ceiling's appearance, and avoiding visual distraction. Most room visitors attributed the sound to the ventilation system. The loudspeakers' volume level could be manually adjusted with the control unit.

The speech condition consisted of level-equalized sentences separated by a 600 ms break between them. The A-weighted equivalent SPL of the speech condition, measured in the listener's position, was 50 dB L_{Aeq}. The level includes the breaks between the sentences and between the words. Successive sentences were mixed to eliminate any coherent plot. The measured spectrum of speech is displayed in Fig. 1. The spectrum shape conforms with the standardized spectrum of speech according to the ISO 3382-3 standard (ISO, 2012).

The sentences were obtained from a stereo audiobook (Jansson, 2015) narrated by a single adult male. The selection process involved two stages: a speech break analysis and a sentence analysis. The speech break analysis employed audio software (Adobe Audition 2020, Adobe Inc., San Jose, California, USA) to identify breaks in the audiobook stream, typically occurring between spoken sentences. A break was defined when the momentary A-weighted SPL (average of both channels) dropped 15 dB below the overall average SPL of the audiobook. Sentence candidates were deemed viable if the break lasted over 200 ms. Each sentence candidate was segmented at cutoff points where the SPL dropped 15 dB below the average level. These candidates were saved as individual wave files, and their A-weighted equivalent SPL was normalized using custom software (MATLAB R2019b, The MathWorks Inc., Natick, Massachusetts, USA). The number of sentence candidates totaled 1749.

The sentence candidate files were reviewed to identify suitable sentences. Sentence candidates were excluded based on criteria such as violating grammatical definitions of main or subordinate clauses, having fewer than 3 words or more than 8 words, containing identifiable names (places or persons), or drawing excessive attention through shouting or laughter. A total of 1303 sentences were deemed suitable.

The final speech file comprised these suitable sentences, arranged in a pseudorandom order, with the condition that consecutive sentences in the original audiobook were not placed consecutively in the final speech file. A 600 ms silence interval was inserted between each sentence. The spectrum of the final speech file was adjusted to conform to the standardized human speech spectrum (ISO, 2012) using custom software in Matlab.

The total duration of the final speech file was 90 min. The first 15 min were only needed in speech condition. The rest of the final speech file was not used in this experiment.

The playback system in the experimental room consisted of a portable computer, a media player (VLC 3.0.11), a sound card (Rubix 22, Roland Co., Hamamatsu, Japan), and two active loudspeakers (8020A, Genelec Ltd., Iisalmi, Finland). The computer was located outside the experimental room to avoid any influence on background noise. The loudspeakers were installed to the wall behind the participant (see Fig. S1 in the Supplementary material). The levels of both sound card and loudspeakers were set to maximum. Adequate playback level was set using the level adjustment of the computer (Windows media player).

The playback levels of both *conditions* were inspected in the experimental room by measuring the SPL in one-third octave bands at a location of typical participant's ears during the experiment (120 cm from the floor). The measurements were made using a head-and-torso simulator (B&K 4100, Brüel & Kjær Sound and Vibration Measurement, Naerum, Denmark), a microphone power supply (B&K 2804, Brüel & Kjær Sound and Vibration Measurement, Naerum, Denmark), a soundcard (D-audio USB Pre-Amp, Duran Audio Ltd., The Netherlands) and a measurement software (B&K Pulse Sound Quality 15.1.0, Brüel & Kjær Sound and Vibration Measurement, Naerum, Denmark).

The speech file was equalized using a 1/3 octave band equalizer to compensate for the effect of the loudspeaker and the room. The spectrum of the SPL was equalized to match the ISO 3382-3 speech spectrum (ISO, 2012) by using a head-and-torso simulator (B&K 4100), a microphone power supply (B&K 2804), a portable multitrack recorder (TASCAM DR-680MKII, Montebello, California, USA), and custom software in Matlab. A frequency-dependent diffuse-field correction was applied to the measurement results to compensate for the increase in SPL caused by the head-and-torso simulator above 800 Hz.

2.6. Dependent measures

The dependent measures are presented in Table 1. They consist of psychological (subjective) measures, performance measures based on two cognitive tasks, and physiological measures.

2.6.1. Psychological measures

Table 1 provides details on variables, questions, response scales, and the phase at which the measures were collected. Subjective experience was assessed after each task using *annoyance* and *workload* ratings on an 11-point scale. *Annoyance* in relation to sound was inquired using a question recommended by ISO/TS 15666 (ISO, 2003b). Since *annoyance* and *workload* were estimated after each task in intermediate questionnaire 1 (IQ1), each condition contained two estimations, and the final score was derived as their average. General evaluations of the sound environment were conducted at the end of the conditions in intermediate questionnaire 2 (IQ2) employing selected variables that describe acoustic satisfaction (Haapakangas et al., 2011).

2.6.2. Performance measures (cognitive tasks)

The visual serial recall task (SR) measures verbal short-term memory. These task specifications are adapted from studies on the performance effects of task irrelevant speech (Haapakangas et al., 2011; Haka et al., 2009; Radun et al., 2021). In this task, participants were presented numbers on a display from one to nine in a random order. The presentation time for each number was one second, and the inter-stimulus interval was 1.5 s. Participants were instructed to remember the order of the numbers and report the order in a 3 × 3 answer window that appeared on the screen 10 s after the presentation of the last number had finished. The response was given by selecting the correct number with a mouse. The response window had a bar showing which number was in order. Once the number was selected, changes to that number were no longer possible. The response window disappeared when the whole series had been reported or when 15 s had passed. The task was repeated 3 times in the practice phase, and 11 times in the test phase of each *condition* (2). The first trial of the series of 11 was excluded from the analysis.

The N-back task requires working memory, more specifically online monitoring, updating, and manipulation of remembered information (Owen et al., 2005). The sequences of letters were presented on a display: one letter at a time, for 500 ms with an inter-stimulus interval of 2500 ms. The instruction was to press “YES” each time the presented letter was the same as that presented *n* letters back. Otherwise, the instruction was to press “NO”. Each level consisted of $30 + n$ letters ($n = 1$, or 3), of which 9 letters (30 %) were target letters requiring a “YES” response. To ensure that each $30 + n$ set had exactly 9 target letters, 8 pseudorandomized lists were used, in which the places of the target letters were predetermined. The non-target letters were randomly selected each time, and the target letter was the same letter as *n* letters back. Whether the letter was presented as a capital or a small letter was also assigned separately each time with a 50 % probability. We applied the 1-back and 3-back tasks. However, the 1-back task was only included in the experiment as it makes the 3-back task easier to understand and the experiment lighter to perform. Therefore, only the responses to 3-back were reported, as 1-back showed performance saturation. The mean accuracy was the proportion of correct answers. Reaction times that deviated by >2.5 standard deviations from a participant's mean were excluded. The reaction times were calculated as the means of the reaction times to correct answers.

For both tasks, accuracy and response time were examined. Response time was included, as some studies have suggested that stress might

Table 1
The description of the variables in the experiment. The phase refers to Section 2.8 and Fig. 2.

Variable name	Description	Response/unit	Range	Phase
Psychological measures - subjective experience				
Annoyance	"How much does the sound annoy, irritate, or bother you?"	0 not at all, 10 extremely	0–10	IQ1
Workload	"How burdensome was performing the previous task in your opinion?"	0 not at all, 10 extremely	0–10	IQ1
Psychological measures - acoustic satisfaction				
Pleasantness	"Sound environment was pleasant."	1 completely disagree - 5 completely agree	1–5	IQ2
Distraction	"Sound environment bothered my concentration."	1 completely disagree - 5 completely agree	1–5	IQ2
Performance impairment	"Sound environment decreased my task performance."	1 completely disagree - 5 completely agree	1–5	IQ2
Efficiency	If you should work daily with similar tasks in a similar sound environment that you just experienced. "I could work efficiently for long times."	1 completely disagree - 5 completely agree	1–5	IQ2
Performance measures				Test phase
Serial recall (SR) accuracy	Remembering 9 numbers presented in random order on display. Proportion of correct answers of ten trials.	Accuracy per position	0–1	
Serial recall (SR) response time	The time until the response was complete from when the response window appeared.	Seconds (s)		
3-back accuracy	The proportion of correct answers from all presented letters.	Mean accuracy	0–1	
3-back reaction time	Reaction time from correct responses.	Seconds (s)		
Physiological measures				Continuously
HR	Heart rate	Beats per minute (bpm)		
HRV HF	Power in high frequency band range 0.15–0.4 Hz in normalized units (HF/total power-VLF)x100	Normalized units (n.u.)		
HRV LF	Power in low frequency band range 0.04–0.15 Hz in normalized units (LF/total power-VLF)x100	Normalized units (n.u.)		
HRV LF/HF	LF/HF ratio			
RMSSD	Root mean square of successive RR differences	Milliseconds (ms)		
Background variables				
Noise sensitivity	21-item noise sensitivity scale		21–126	Beginning questionnaire
WM capacity	Operation span task		0–1	

WM = working memory, IQ = intermediate questionnaire, HR = heart rate, HRV = heart rate variability, HF = high frequency, LF = low frequency, VLF = very low frequency, RR = time between heart beats

cause a strategic change, e.g., a possible speed-accuracy trade-off, when the performance is faster, but less accurate during the stressor (Hockey, 1997).

2.6.3. Physiological measures

Heart rate (HR) and heart rate variability (HRV) were measured with an ECG sensor (Faros 180, Bittium Biosignals Ltd., Finland). Participants were instructed to attach a textile belt and Stingray adapter under the chest muscle line. The accelerometer sampling rate was 25 Hz, and the dynamic range was ±4 g. The ECG sampling rate was 250 Hz. R-R interval data were analyzed using dedicated software (Kubios HRV, Kubios Ltd., Finland). The sensor was synchronized with the computer running the experimental tasks, which marked the exact duration of the *conditions*. These markers were later entered into the ECG data.

Frequency-domain measures are preferred for short-term measurements as they reflect the balance between the sympathetic and parasympathetic autonomic nervous systems, which are related to stress (Kim et al., 2018). Frequency-domain analyses were performed on the FFT spectrum to determine the powers of *LF* (low frequency, 0.04–0.15 Hz) and *HF* (high frequency, 0.15–0.4 Hz) bands in normalized units (n.u.) as well as the ratio of the LF and HF band powers (*HRV LF/HF*) (Tarvainen et al., 2014). Additionally, alongside the frequency-domain analysis, a time-domain analysis of HR and HRV were examined. The activity of the parasympathetic nervous system is evident in the time-domain analysis variables, as they reflect beat-to-beat changes (Kim et al., 2018). For the time-domain analysis of HRV, the root mean square of the successive R-R interval differences (*RMSSD*) was determined. This parameter describes the difference between R-R intervals; a large difference signifies more variability, thus less stress, while a small difference indicates less variability and more stress (Tarvainen et al., 2014).

The data from one participant was excluded due to missing data (loose HR belt). Gender is suggested to influence short-term HRV measures, especially in the relatively young age group used in our study (Voss et al., 2015). Therefore, gender was included as a covariate to the HR and HRV analyses. The analysis was based on condition-based and not task-based analysis as with task-based analysis there were notably more outliers, possibly due to short samples (one condition was on average 13 min and the duration of the n-back task <5 min).

2.7. Background variables – the uniformity of the presentation order groups

To ensure the uniformity of the *presentation order* groups, the variables that might influence performance during speech, such as working memory (Hughes et al., 2013; Sörqvist, 2010) and noise sensitivity (Belojević et al., 1992; Ellermeier and Zimmer, 1997) were measured. These background variables are depicted in Table 1 and below.

Noise sensitivity was estimated using Weinstein's 21-item questionnaire (Weinstein, 1978).

Working memory (WM) capacity was measured using a modified operation span task (Turner and Engle, 1989). In this task, equation-word pairs were presented sequentially. The task involved indicating the correctness of the equations and memorizing the presented words.

The equations consisted of a multiplication and addition or subtraction calculations; for example $3 \times 9 + 8 = 35$. Participants had 10 s to determine the equation's correctness and select the appropriate option from the display using a mouse.

Following each equation, a word was presented on a display for 2 s. Short and common words were selected from the Word Mill lexical search program (Laine and Virtanen, 1999) based on the criteria: 5–7 letters, 2–3 syllables per word, and a frequency range of 50–999. The inter-stimulus interval was 0.5 s. Participants were instructed to verbally articulate the word and remember it.

The number of words and equations in each set ranged from 3 to 7, ensuring that each set was presented twice. The words and equations were randomly selected from word and equation lists, ensuring that each equation and word appeared only once in the experimental session. Once all the words and equations in a set were presented, a free text box appeared where participants wrote down as many words as they could remember from the last set. Participants were informed that the order of the remembered words was not relevant.

To maintain the participants' attention on both tasks, they were instructed to strive for a minimum accuracy of 85 % on equations in each set. Following each set, participants received feedback on their accuracy with equations. Prior to the actual task, participants were given instructions and a practice session. The practice session comprised sets consisting of 2, 3, and 4 equations and words, distinct from the words and equations of the test phase.

WM capacity was determined by the proportion of correctly remembered words from all words presented (using partial-credit load scoring, Conway et al., 2005). The words could be written in any order, and wrong words were not calculated. Minor misspellings were accepted when unambiguous.

2.8. Procedure

The procedure is outlined in Fig. 2. The participants were tested one by one in the same experimental room.

In the preparation phase, participants signed the informed consent form, put on the HR monitor and their hearing was tested with a Screening Audiometer (Madsen Micromate 304, Otometrics). The beginning questionnaire included questions about the participant's current state, such as sleep duration the previous night and their current feelings, along with the *noise sensitivity* questionnaire.

In the WM capacity phase, the experiment leader explained the operation span task, and participants practiced it. Subsequently, participants performed the task alone in the experimental room.

During the practice phase, participants practiced the serial recall task and N-back tasks.

Following this, the test phases 1 and 2 were conducted. Both phases included the serial recall task, the IQ1, N-back task, IQ1, and IQ2. However, the test phases had different sound exposures (*conditions*). The *presentation order* of the *conditions* was balanced across the participants so that one group of participants started with the quiet condition in Test phase 1 (Quiet first group) and the other with the Speech condition (Speech first group). In all other experiment phases except the test phases, the background noise conformed with the quiet condition (no exposure).

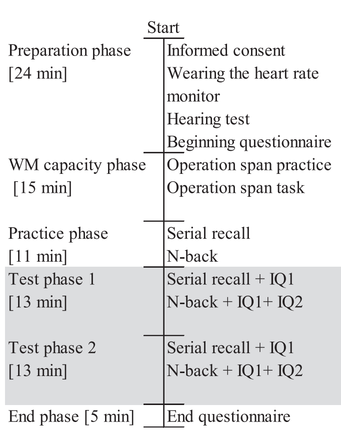

Fig. 2. Experimental procedure contained six phases. Mean durations are indicated in brackets. The *conditions* were presented in the Test phases 1 and 2, so that for participants with uneven participant numbers, the quiet *condition* was in Test phase 1 (Quiet first group) and for participants with even participant numbers, the speech *condition* was in Test phase 1 (Speech first group). IQ denotes intermediate questionnaire.

In the end phase, participants completed the end questionnaire, the heart rate monitor was removed, and the participant fee was provided.

On average, the whole experiment took 1 h 20 min to complete, with each test phase lasting on average 13 min.

The experiment was implemented using MATLAB R2015a with Psychtoolbox – 3 (PTB; psychtoolbox.org) (Brainard, 1997). Only the beginning and end questionnaires were presented using an internet survey tool Webropol (Webropol Ltd., Helsinki, Finland).

2.9. Statistical analysis

The analysis was conducted using IBM SPSS Statistics for Windows, version 28 (IBM Corp., Armonk, NY, USA) and the significance level was set at $p < 0.05$.

Initially, we aimed to ensure that the *presentation order* groups did not exhibit differences in variables that might impact performance during speech, namely working memory (Hughes et al., 2013; Sörqvist, 2010) and noise sensitivity (Belojević et al., 1992; Ellermeier and Zimmer, 1997). Differences in *WM capacity* and *noise sensitivity* between *presentation order* groups were examined using Analysis of Variance (ANOVA).

Subsequently, all dependent variables underwent assessment for normality and outliers using the Kolmogorov-Smirnov test of normality and by examining skewness and kurtosis values (±2 was used as the cutoff criteria). Variables meeting either of these criteria were considered sufficiently normal for mixed-models analysis of variance (mANOVA). Variables failing to meet the criteria were scrutinized for outliers, which were potentially removed. One participant was removed in *HRV LF/HF* analysis. Additionally, one participant was removed from the results of 3-back task (both *3-back accuracy* and *3-back response time*) as the results indicated a misunderstanding of the task logic.

For variables meeting these criteria, mANOVA was employed, utilizing *condition* (2) as the within-subject variable and *presentation order* (2) as a between-subjects variable. If the interaction between the *condition* and the *presentation order* was significant, pairwise comparisons between the *conditions* in different *presentation order* groups were examined. For physiological variables, *gender* was included as a covariate.

The variables related to the direct estimation of sound environment effects on experience or performance (*annoyance, distraction, performance impairment*) did not satisfy the normality requirements, which was expected given the ideal conditions during quiet. These variables were examined using the Wilcoxon signed rank test as it does not require a normal distribution. The effect size was determined using $r = z/N^{1/2}$. However, the Wilcoxon signed rank test exclusively examines the differences between repeated measures (within-subject variable), and the effect of *presentation order* (between-subject variable) could not be tested with it. For assessing the between-subjects effects of *presentation order*, the difference between the ratings in the speech and quiet conditions (speech minus quiet) was computed. This difference variable adhered to the normality requirements, enabling the use of analysis of variance. Since the main effects of *condition* and *presentation order* were evaluated separately, no interaction could be determined for these variables. The information on the interactions was considered important only for the variables related to extra strain, such as *workload* and *efficiency*.

3. Results

3.1. The uniformity of presentation order groups

The *presentation order* groups did not differ in the background variables *WM capacity* ($F(1, 28) = 0.154, p = 0.698, \eta_p^2 = 0.005$) nor in *noise sensitivity* ($F(1, 28) = 2.8, p = 0.106, \eta_p^2 = 0.091$). Hence, these variables were not included in the subsequent analyses.

3.2. Effect of condition

3.2.1. Psychological measures

The results of the psychological measures are presented in Table 2.

Subjective experience was better, when working during quiet than during speech. *Annoyance* (Z = -4.54, $p < 0.001$, $r = -0.83$) and subjective *workload* ($F(1, 28) = 7.3$, $p = 0.011$, $\eta_p^2 = 0.208$) were higher during speech compared to quiet. In addition, acoustic satisfaction was higher during quiet compared to speech. *Pleasantness* was higher during quiet than during speech ($F(1, 28) = 36.0$, $p < 0.001$, $\eta_p^2 = 0.562$) as was *efficiency* ($F(1, 28) = 47.6$, $p < 0.001$, $\eta_p^2 = 0.630$). *Distraction* (Z = -4.58, $p < 0.001$, $r = -0.84$), and *performance impairment* (Z = -4.35, $p < 0.001$, $r = -0.79$) were lower during quiet compared to speech (Table 2).

3.2.2. Performance measures

The results of the performance measures are presented in Table 3.

The SR accuracy was lower during speech compared to quiet ($F(1, 28) = 20.50$, $p < 0.001$, $\eta_p^2 = 0.423$) (Fig. 3, Table 3). This was the only performance measure that had a significant main effect related to *condition.* The main effect of condition was non-significant for *SR response time* ($F(1, 28) = 2.74$, $p = 0.109$, $\eta_p^2 = 0.089$), *3-back accuracy* ($F(1, 27) = 1.0$, $p = 0.321$, $\eta_p^2 = 0.037$), and *3-back reaction time* ($F(1, 27) = 0.4$, $p = 0.524$, $\eta_p^2 = 0.015$).

3.2.3. Physiological measures

The results of the physiological measures are presented in Table 4.

HR was faster during speech than during quiet ($F(1, 26) = 5.9$, $p = 0.022$, $\eta_p^2 = 0.185$) (Fig. 4a). The frequency-domain analysis of HRV indicated higher stress during speech in comparison to quiet. *HRV LF* was higher during speech than during quiet ($F(1, 26) = 9.9$, $p = 0.004$, $\eta_p^2 = 0.276$) (Fig. 4b). *HRV HF* was lower during speech than quiet ($F(1, 26) = 9.9$, $p = 0.004$, $\eta_p^2 = 0.275$) (Fig. 4c). *HRV LF/HF* was higher during speech than quiet ($F(1, 25) = 8.1$, $p = 0.009$, $\eta_p^2 = 0.245$) (Fig. 4d). The time-domain measure *RMSSD* did not exhibit a significant main effect of the *condition* ($F(1, 26) = 0.7$, $p = 0.420$, $\eta_p^2 = 0.025$).

3.3. Effect of presentation order

3.3.1. Psychological measures

The *presentation order* influenced neither *annoyance* ($F(1, 28) = 0.125$, $p = 0.726$, $\eta_p^2 = 0.004$) nor subjective *workload* ($F(1, 28) = 1.4$, $p = 0.243$, $\eta_p^2 = 0.048$). *Workload* showed no significant interaction between the *condition* and the *presentation order* ($F(1, 28) = 2.6$, $p = 0.120$, $\eta_p^2 = 0.084$).

The presentation order did not influence measures related to acoustic satisfaction (*pleasantness*: $F(1, 28) = 2.3$, $p = 0.140$, $\eta_p^2 = 0.076$; *efficiency*: $F(1, 28) = 2.7$, $p = 0.116$, $\eta_p^2 = 0.086$; *distraction*: $F(1, 28) = 0.017$, $p = 0.897$, $\eta_p^2 = 0.001$; *performance impairment*: $F(1, 28) = 0.014$, $p = 0.907$, $\eta_p^2 = 0.000$) nor was there an interaction between the *condition* and the *presentation order (pleasantness*: $F(1, 28) = 0.4$, $p = 0.510$, $\eta_p^2 = 0.016$; *efficiency:* $F(1, 28) = 0.0$, $p = 0.911$, $\eta_p^2 < 0.001$). The interaction was not examined for *annoyance*, *distraction*, and *performance impairment* as these variables were non-normally distributed.

In conclusion, the psychological measures were neither influenced by the *presentation order*, nor was there an interaction between the *condition* and the *presentation order* (Table 2).

3.3.2. Performance measures

SR accuracy was not influenced by the *presentation order* ($F(1, 28) = 0.43$, $p = 0.517$, $\eta_p^2 = 0.015$), nor was there an interaction between the *condition* and *presentation order* ($F(1, 28) = 3.9$, $p = 0.059$, $\eta_p^2 = 0.121$).

For *SR response time,* the main effect of the *presentation order* was not significant ($F(1, 28) = 0.0$, $p = 0.841$, $\eta_p^2 = 0.001$). Nevertheless, the *SR response time* exhibited an interaction between the *presentation order* and the *condition* ($F(1, 28) = 5.05$, $p = 0.033$, $\eta_p^2 = 0.153$) (Fig. 5a). When speech was first, the *SR response time* was prolonged during speech (M = 11.1 s, SD = 0.5) compared to quiet (M = 10.2 s, SD = 0.4) ($F(1, 28) = 7.6$, $p = 0.010$, $\eta_p^2 = 0.214$). Conversely, when quiet was presented first, no difference existed between the conditions ($F(1, 28) = 0.2$, $p = 0.678$, $\eta_p^2 = 0.006$).

No main effect of the *presentation order* was observed for the *3-back accuracy* ($F(1, 27) = 1.1$, $p = 0.313$, $\eta_p^2 = 0.038$). However, *3-back accuracy* displayed an interaction between the *condition* and the *presentation order* ($F(1, 27) = 13.2$, $p = 0.001$, $\eta_p^2 = 0.328$) (Fig. 5b). When speech was presented first, accuracy during speech was lower (M = 0.762, SD = 0.03) compared to quiet (M = 0.847, SD = 0.02) ($F(1, 27) = 11.1$, $p = 0.002$, $\eta_p^2 = 0.292$). Yet, when quiet was presented first, the *conditions* did not differ from each other in *3-back accuracy* ($F(1, 27) = 3.3$, $p = 0.080$, $\eta_p^2 = 0.109$).

The main effect of *presentation order* was non-significant for the *3-back reaction time* ($F(1, 27) = 0.5$, $p = 0.492$, $\eta_p^2 = 0.018$). However, an interaction between *condition* and *presentation order* was evident ($F(1, 27) = 8.3$, $p = 0.008$, $\eta_p^2 = 0.234$) (Fig. 5c). When quiet was presented first, reaction time was longer during quiet (M = 1.00 s, SD = 0.08 s) compared to speech (M = 0.92 s, SD = 0.08 s) ($F(1, 27) = 6.0$, $p = 0.021$, $\eta_p^2 = 0.181$), yet the *conditions* did not differ when speech was presented first ($F(1, 27) = 2.6$, $p = 0.121$, $\eta_p^2 = 0.087$).

In conclusion, for performance measures, the *presentation order* did not show a main effect, but an interaction between the *condition* and the *presentation order* was evident for measures not showing the effect of the *condition* (Table 3).

Table 2
The means, the standard deviations (SD), and the inference statistics for the main effects of the *condition,* and the *presentation order* as well as their interactions are presented for the psychological measures. The bolded values denote the significant effects ($p < 0.05$). The interaction between the *condition* and the *presentation order* is missing (–) from the variables that were not tested with mANOVA due to their non-normal distribution. The response scales are given under the table.

Variables	Descriptive statistics		Inference statistics		
	Quiet	Speech	*Condition*	*Presentation order*	*Condition* × *presentation order*
	Mean (SD)	Mean (SD)	*p*-value	*p*-value	*p*-value
Psychological measures – subjective experience					
Annoyance[a]	**1.0 (1.9)**	**5.4 (3.1)**	**<0.001**	0.726	–
Workload[a]	**5.6 (2.2)**	**6.5 (2.3)**	**0.011**	0.243	0.120
Psychological measures – acoustic satisfaction					
Pleasantness[b]	**4.3 (1.1)**	**2.5 (1.3)**	**<0.001**	0.140	0.510
Distraction[b]	**1.4 (0.9)**	**3.8 (1.3)**	**<0.001**	0.897	–
Performance impairment[b]	**1.3 (0.6)**	**3.3 (1.3)**	**<0.001**	0.907	–
Efficiency[b]	**3.9 (1.2)**	**1.9 (1.2)**	**<0.001**	0.116	0.911

[a] 0 not at all–10 extremely.

[b] 1 completely disagree–5 completely agree.

Table 3
The means, the standard deviations (SD), and the inference statistics for the main effects of the *condition*, and the *presentation order* as well as their interactions are presented for the performance measures. The bolded values denote the significant effects ($p < 0.05$).

Variables	Descriptive statistics		Inference statistics		
	Quiet	Speech	*Condition*	*Presentation order*	*Condition* × *Presentation order*
	Mean (SD)	Mean (SD)	*p*-value	*p*-value	*p*-value
SR accuracy	**0.61 (0.14)**	**0.51 (0.14)**	**<0.001**	0.517	0.059
SR response time [s]	10.4 (1.60)	10.8 (2.00)	0.109	0.841	**0.033**
3-back accuracy	0.83 (0.07)	0.81 (0.11)	0.321	0.313	**0.001**
3-back response time [s]	0.93 (0.29)	0.92 (0.25)	0.524	0.492	**0.008**

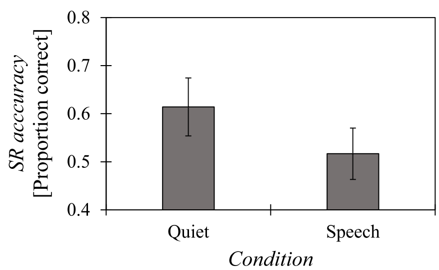

Fig. 3. The mean values of the performance measure with the significant main effect of the *condition*. Error bars denote the 95 % confidence interval.

3.3.3. Physiological measures

The *presentation order* had no main effect on *HR* ($F(1, 26) = 0.0$, $p = 0.876$, $\eta_p^2 = 0.001$), but an interaction between the *condition* and the *presentation order* was observed ($F(1, 26) = 35.8$, $p < 0.001$, $\eta_p^2 = 0.580$) (Fig. 6a). When speech was presented first, *HR* was faster during speech ($M = 81.3$ bpm, SD = 13.5) compared to quiet ($M = 77.8$ bpm, SD = 12.6) ($F(1, 26) = 36.9$, $p < 0.001$, $\eta_p^2 = 0.586$). Conversely, when quiet was presented first, *HR* was faster during quiet ($M = 80.6$ bpm, SD = 10.4) than during speech ($M = 79.0$ bpm, SD = 11.6) ($F(1, 26) = 6.2$, $p = 0.020$, $\eta_p^2 = 0.192$).

The presentation order showed no significant main effect on the frequency-domain variables (*HRV LF:* $F(1, 26) = 0.3$, $p = 0.612$, $\eta_p^2 = 0.010$; *HRV HF:* $F(1, 26) = 0.3$, $p = 0.613$, $\eta_p^2 = 0.010$; *HRV LF/HF*: $F(1, 25) = 0.3$, $p = 0.752$, $\eta_p^2 = 0.004$). Furthermore, no interaction between the *condition* and the *presentation order* was observed (*HRV LF*: $F(1, 26) = 0.6$, $p = 0.460$, $\eta_p^2 = 0.021$; *HRV HF*: $F(1, 26) = 0.6$, $p = 0.463$, $\eta_p^2 = 0.021$; *HRV LF/HF:* $F(1, 25) = 2.3$, $p = 0.143$, $\eta_p^2 = 0.084$).

The time-domain measure *RMSSD* did not exhibit a significant main effect of the *presentation order* ($F(1, 26) = 0.0$, $p = 0.885$, $\eta_p^2 = 0.001$). However, an interaction between the *condition* and the *presentation order* was observed ($F(1, 26) = 12.3$, $p = 0.002$, $\eta_p^2 = 0.322$) (Fig. 6b). When speech was presented first, RMSSD was lower, indicating reduced variability between successive heart beats during speech ($M = 42.7$ ms, SD = 8.3 ms) compared to quiet ($M = 46.2$ ms, SD = 9.3 ms) ($F(1, 26) = 12.8$, $p = 0.001$, $\eta_p^2 = 0.329$). However, when quiet was presented first, no difference between the *conditions* was found ($F(1, 26) = 2.1$, $p = 0.158$, $\eta_p^2 = 0.075$).

In conclusion, the frequency-domain measures showed no influence for the *presentation order*, whereas an interaction of the *presentation order* and the *condition* was visible in the time-domain measures of HRV (Table 4).

Fig. 7 presents a summary of the results concerning the main effect of the *condition* and the interaction between the *condition* and the *presentation order.* The figure indicates the significant differences and the direction of the difference. The column concerning the main effects of the *presentation order* was omitted, as it did not exhibit significance in relation to any variable.

4. Discussion

4.1. Effect of condition

Our study showed, that compared to working during quiet, working during task-irrelevant speech showed a reduced accuracy of the verbal serial recall task, higher annoyance and workload, and a less pleasant sound environment. Furthermore, participants rated speech as disturbing concentration, impairing performance as well as hampering efficient working. These results are consistent with previous studies investigating the impact of speech on performance (Haapakangas et al., 2020; Liebl et al., 2016; Schlittmeier et al., 2008) and stress (Evans and Johnson, 2000; Radun et al., 2021; Tafalla and Evans, 1997).

Physiologically, working during speech resulted in lower HRV HF, and higher HR, HRV LF, and HRV LF/HF ratio in comparison to working during quiet. These measures indicate changes in the autonomic balance, namely higher sympathetic activity, and reduced parasympathetic activity. These changes can be interpreted as indicators of an elevated physiological stress level. Hence, the frequency-domain analysis of HRV illustrated an additional strain associated with working during irrelevant speech, as opposed to working in a quiet environment. This finding demonstrates that the additional strain is not solely related to performing tasks compared to rest, which might have been a possible interpretation of a previous study (Kristiansen et al., 2009).

Table 4
The means, the standard deviations (SD), and the inference statistics for the main effects of the *condition,* and the *presentation order* as well as their interactions are presented for the physiological measures based on ECG data. The bolded values denote the significant relations ($p < 0.05$).

Variables	Descriptive statistics		Inference statistics		
	Quiet	Speech	*Condition*	*Presentation order*	*Condition* × *Presentation order*
	Mean (SD)	Mean (SD)	*p*-value	*p*-value	*p*-value
HR [bpm]	**79.1 (11.5)**	**80.2 (12.4)**	**0.022**	0.876	**<0.001**
HRV LF [n.u.]	**66.7 (12.8)**	**69.1 (13.4)**	**0.004**	0.612	0.460
HRV HF [n.u.]	**33.2 (12.8)**	**30.9 (13.4)**	**0.004**	0.613	0.463
HRV LF/HF	**2.32 (1.3)**	**2.67 (1.40)**	**0.009**	0.752	0.143
RMSSD [ms]	47.3 (33.3)	45.5 (31.2)	0.420	0.885	**0.002**

n.u. = normalized units.

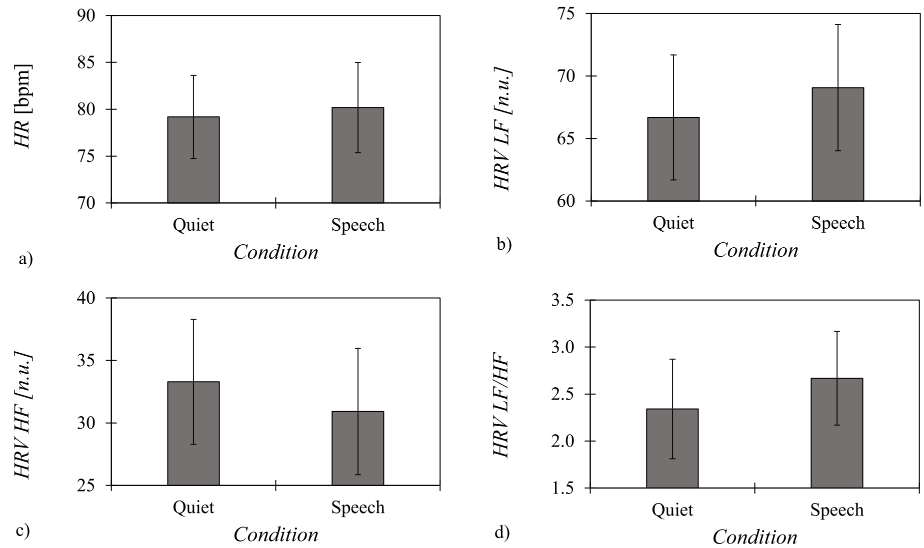

Fig. 4. The mean values of the physiological measures with the significant main effect of the *condition*. Error bars denote the 95 % confidence interval.

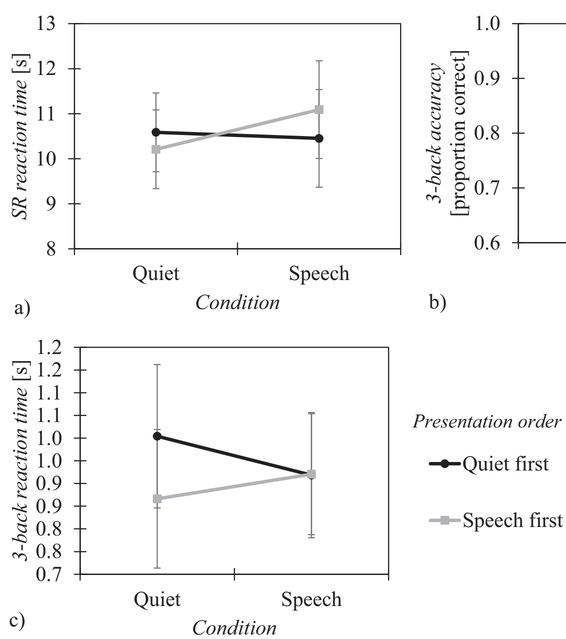

Fig. 5. The mean values of the performance measures that showed a significant interaction between the *condition* and the *presentation order*. Error bars denote the 95 % confidence interval.

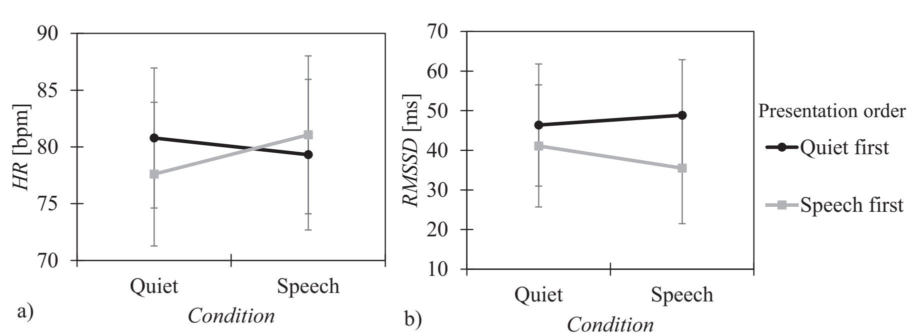

Fig. 6. The mean values of the physiological measures that showed a significant interaction between *condition* and *presentation order*. Error bars denote the 95 % confidence interval.

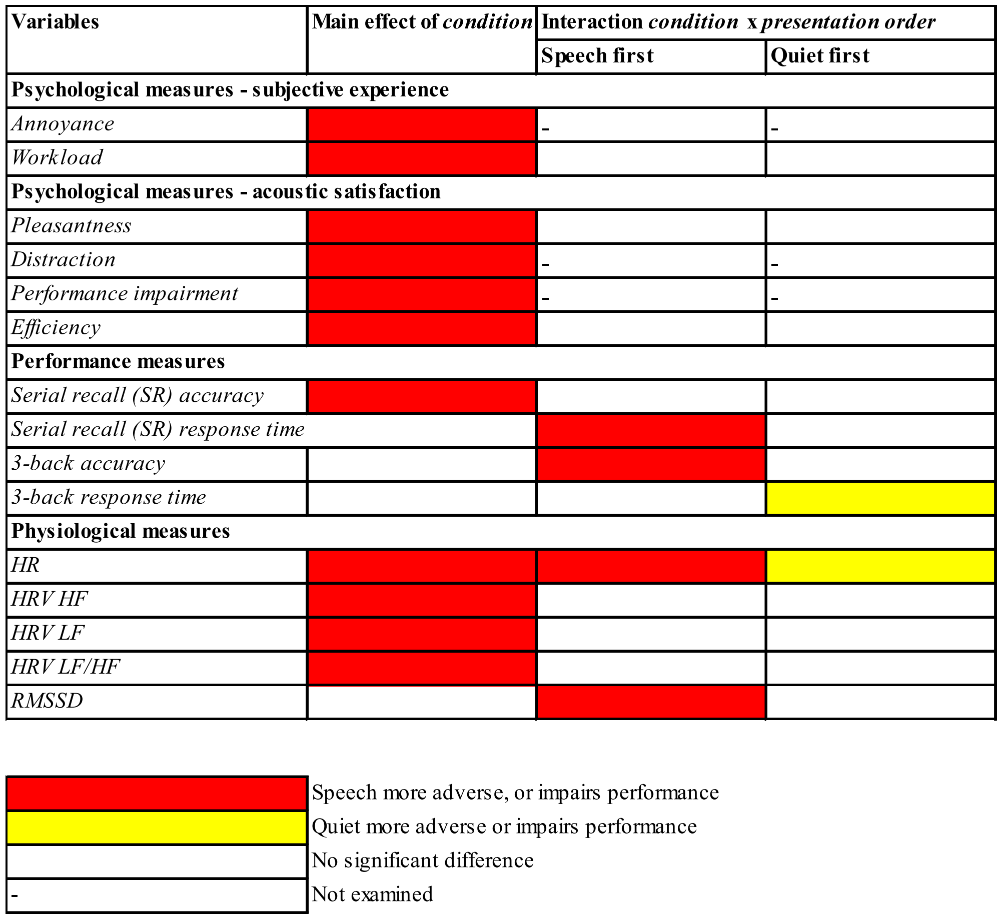

Fig. 7. This summary of results shows the significant effects of the *condition*, and the interaction between the *condition* and the *presentation order*, as well as the direction of the effect for all dependent variables.

Our study demonstrates that HRV is a suitable stress measure for stress related to effects of task-irrelevant sounds, at least during short-term exposure in laboratory studies. Our investigation revealed two key findings: first, the prevalent performance decline associated with irrelevant sound, evidenced by reduced accuracy in the serial recall task, and second, alterations in HRV frequency-domain analysis that suggest heightened stress levels. Notably, HRV frequency-domain analysis is well-suited for assessing short-term stress-related changes (Kim et al., 2018).

Interestingly, the results from our investigation revealed no significant main effect of the *condition* on the 3-back task, neither in terms of accuracy nor reaction times. However, interaction effects emerged, indicating that when speech was presented first, 3-back accuracy was lower during speech compared to quiet; yet, no such distinction between conditions was observed when quiet was presented first. This observation aligns with a prior study that demonstrated a reduction in 3-back accuracy when exposed to speech and steady-state noise compared to a quiet condition (Radun et al., 2021). Notably, the previous study presented the 3-back task twice and found that sound conditions affected accuracy during the first repetition but not the second (Radun et al., 2021). Their finding resonates with our results, also suggesting some transient influence of straining sound conditions.

In our study, irrelevant speech was not associated with longer reaction times in the n-back task as in a previous study (Haapakangas et al., 2014). Their finding was that clearer speech increased reaction times in the n-back task. In contrast to their findings, the present study's results indicate that when quiet was presented as the first condition, reaction times were longer during quiet than during speech. Conversely, when speech was presented first, no discernible difference in reaction times between the conditions was evident. This was an unexpected result as this would indicate slower processing during quiet as the first condition compared to speech as the first condition. Our result might also indicate there was a speed-accuracy trade-off if speech was presented first (Hockey, 1997). The speech content and presentation type could explain at least part of the contradiction with Haapakangas et al. (2014). Our speech signal was selected not to cause much attentional capture according to the duplex-mechanism account of auditory distraction (Hughes et al., 2007), as it was continuous, the breaks between sentences were minimal, and speech was presented always from the same spatial position with the same male voice. Opposite to that, Haapakangas et al. (2014) had four different spatial positions from which speech was presented with randomized order and random breaks between the sentences. Their speech was designed to resemble a situation where four people were talking on the phone in different corners of the room. Their speech might have been more attention-grabbing than ours, and speech interference in our study might be more related to interference-by-process. The N-back task may be influenced by sound conditions, especially if they are novel, but some aspects of the performance can be improved with time or practice. This topic and the cognitive processes behind the performance need more research.

4.2. Effect of presentation order

Our study did not reveal a significant main effect of the *presentation order* on the examined performance, psychological, or physiological variables. This absence of a discernible effect aligns with our initial assumptions. Similarly, an earlier study reported no main effect of the presentation order on verbal short-term memory performance (Farley et al., 2007). However, in their within-subject design, they did observe an interaction, wherein performance was superior when the quiet condition preceded the speech condition, compared to the reverse order (Farley et al., 2007). We also found an interaction between the *presentation order* and the *condition* in certain variables, but not with the serial recall task performance. Most findings indicate that the effects of speech were more pronounced when speech was presented first. *3-back accuracy* during speech was lower than during quiet when speech was the first *condition*, but not when quiet was the first *condition*. Similarly, response times in the serial recall task during speech were slower than during quiet when speech was presented first, but there was no difference between the *conditions* when speech was presented second. However, when quiet was presented first, the *3-back response times* were slower during quiet than during speech, while there was no difference with the opposite *presentation order*. This faster performance and lower accuracy in the 3-back task when speech is the first condition might indicate a possible strategic change called the speed-accuracy trade-off, which is suggested to be one possible strategical change for performance protection under stress (Hockey, 1997).

The interaction of the *presentation order* and the *condition* showed that participants could adapt to specific aspects of the *conditions* or tasks over time. This adjustment appears to be distinct from the irrelevant sound effect observed in the reduced accuracy of the serial recall task. Exploring this phenomenon, a study investigating the serial recall task across repeated sessions involving speech, speech-simulated noise, and quiet environments found that participants did not attain their optimal performance level during the initial session but improved across subsequent sessions (Hellbrück et al., 1996). However, the noise effect remained qualitatively the same across sessions. Therefore, the first session was always performed worse than the second and third, no matter the sound condition, but the impact of the sound condition on performance was the same within the sessions. They concluded that irrelevant speech cannot be habituated to and that during the first sessions, participants were still unpracticed in the task itself. On the other hand, some irrelevant sound studies have shown a habituation effect, which has been linked to complex distractor material. e.g., speech and musical melody (Röer et al., 2014). The habituation could also be seen if the distractor sounds were played only during the maintenance phase of the task. Therefore, the habituation was attributed to attention capture of complex auditory distractors (Röer et al., 2014). This would suggest that 3-back accuracy and the serial recall response times were impaired during speech due to attentional capture related to novel semantically meaningful irrelevant speech. However, if the attention capture of irrelevant speech was the only reason, this effect should have also been visible when speech was presented as the second condition, as speech was a novel sound condition also then. It is possible that some practice effects also influenced performance. Together, little practice and attention capture might load the information processing capacity making speech as the first *condition* a slightly more difficult condition than when it is presented as the second *condition*.

This small extra strain due to irrelevant speech as the first *condition* was also seen in the interactions in physiological measures that were based on the time-domain analysis of HRV. These reflect the beat-to-beat changes in HRV (Kim et al., 2018). When speech was presented first, *HR* was faster during speech than during quiet, but when quiet was presented first, *HR* was slightly faster during quiet than during speech. However, the effect of *presentation order* was not significant, as *HR* was significantly faster during speech than during quiet. HRV *RMSSD* was also lower during speech than during quiet when speech was presented first, but not when speech was presented second. Lower *RMSSD* indicates less variability between the adjacent RR intervals, which indicates higher stress (Kim et al., 2018). Future research could examine the use of these time-domain parameters of HRV in possible monitoring of a stress reaction to deviant and changing-state sounds, as has been done in an irrelevant sound study, utilizing pupillometry in a serial recall task (Marois et al., 2019). There, the changing-state effect was related to tonic increase in pupil size, while the attention deviation was related to was related to pupillary dilation response (Marois et al., 2019).

In summary, when the more challenging condition, i.e., here speech, is the first condition, the physiological measures might indicate the increased effort needed to perform the tasks. Alternatively, the physiological measures might reflect increased stress due to errors or difficulty performing the task. These can also be found in minor additional decreases in performance in 3-back accuracy and serial recall response time. However, the interaction effect was not visible in serial recall task accuracy or frequency-domain analysis of HRV, which showed the main effect of the irrelevant sound. It can be that the minor additional challenge of a new task and sound condition together are needed to show these extra effects.

4.3. Strengths and limitations

Our study had several strengths compared to other studies examining the effects of speech on HRV. First, the *condition* was our within-subject variable. The within-subject approach is, from the perspective of statistical testing, more powerful than the between-subjects approach applied in previous studies (Kristiansen et al., 2009; Radun et al., 2021), especially due to large inter-individual variations in HRV. Second, earlier studies on the effects of HRV and noise have had limited power, especially for a between-subjects design (Radun et al., 2021; Sim et al., 2015). In our study, the number of participants was based on power calculations to ensure sufficient power to detect a medium-sized effect of the *condition*. In the abovementioned studies, the number of participants was not based on such calculations, and the number of participants was deemed to be too small. Third, confounding factors such as *noise sensitivity* and *WM capacity* presumed to possibly affect performance during speech were controlled. Fourth, our speech condition was designed to resemble office conditions. Short sentences with no plot to follow represented hearing, e.g., half of a phone call. In addition, the SPL of the speech condition (50 dB L_{Aeq}) was chosen to fall within the range of moderate SPLs (48–59 dB L_{Aeq}) that typically take place in occupied open-plan offices (Yadav et al., 2021). Therefore, our findings can be applied to understand the effects of irrelevant speech in workplaces.

A limitation is that our power calculations were made with the assumptions that there are no between-subjects effects. This was true, but to reliably detect the possible interaction effects, especially if these were expected to be small, the sample size is too small, and further studies with more participants are needed. Therefore, the results regarding the interaction of time-domain measures of HRV are weak and need further verification. Furthermore, we only had two conditions and repetitions of each task. More repetitions would be interesting to examine the practice or habituation effects. In addition, our material was designed to be ecologically valid for an office, and it had semantically meaningful sentences. Thus, it was not purely related to changing-state or deviant processing as for example presenting letter lists can be (Hughes et al., 2007). We aimed to make material that would interfere with the serial processing, but attentional deviation due to semantics cannot be ruled out. In the future, it would be interesting to examine the stress related to interference-by-process and attentional deviation with HRV with an experiment that could clearly separate the mechanisms. In addition, a more ecologically valid field experiment examining HRV with controlled tasks in a natural environment would also be intriguing.

5. Conclusions

Irrelevant speech presented at a moderate level (50 dB L_{Aeq}) impaired performance, worsened experience, and increased physiological stress compared to the quiet condition without speech (35 dB L_{Aeq}). The stress effects related to irrelevant speech were seen in the frequency-domain measures of HRV. Our study proves that HRV (which is based on ECG analysis) is an easy and non-invasive stress measure that is usable, especially in laboratory experiments on sound effects. The largest influence of speech was found when it was presented first, followed by the quiet condition. This finding indicates an additional interaction effect of working during speech. The load of speech is more pronounced if the tasks are not well-practiced and sound conditions are new. This extra load was also visible as an additional performance challenge and higher stress, as indicated by time-domain measures of HRV. These effects were reduced when the task-irrelevant speech was presented as the second sound condition. However, overall, working during speech is clearly stressful, reduces performance, and causes annoyance, workload, and a less pleasant sound environment, which cannot be corrected by practice or habituation. Therefore, the exposure to irrelevant speech should be minimized in workplaces where people must perform cognitively demanding tasks.

CRediT authorship contribution statement

Jenni Radun: Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft. **Henna Maula:** Data curation, Formal analysis, Methodology, Software, Writing – review & editing. **Iida-Kaisa Tervahartiala:** Investigation, Writing – review & editing. **Ville Rajala:** Methodology, Software, Writing – review & editing. **Sabine Schlittmeier:** Formal analysis, Writing – review & editing. **Valtteri Hongisto:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Acknowledgements

This experiment was made in ActiveWorkspace, a project funded by the Academy of Finland [decision number 314597].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijpsycho.2024.112352.

References

Belojević, G., Öhrström, E., Rylander, R., 1992. Effects of noise on mental performance with regard to subjective noise sensitivity. Int. Arch. Occup. Environ. Health. https://doi.org/10.1007/BF00378288.

Brainard, D.H., 1997. The psychophysics toolbox. Spat. Vis. 10, 433–436. https://doi.org/10.1163/156856897X00357.

Brysbaert, M., 2019. How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. J. Cogn. 2 (1), 1–38. https://doi.org/10.5334/joc.72.

Conway, A.R.A., Kane, M.J., Bunting, M.F., Hambrick, D.Z., Wilhelm, O., Engle, R.W., 2005. Working memory span tasks: a methodological review and user's guide. Psychon. Bull. Rev. https://doi.org/10.3758/BF03196772.

Ellermeier, W., & Zimmer, K. (1997). Individual differences in susceptibility to the "irrelevant speech effect." J. Acoust. Soc. Am., 102(4), 2191–2199. doi:https://doi.org/10.1121/1.419596.

Evans, G.W., Johnson, D., 2000. Stress and open-office noise. J. Appl. Psychol. 85 (5), 779–783. https://doi.org/10.1037/0021-9010.85.5.779.

Farley, L.A., Neath, I., Allbritton, D.W., Surprenant, A.M., 2007. Irrelevant speech effects and sequence learning. Mem. Cogn. 35 (1), 156–165. https://doi.org/10.3758/BF03195951.

Haapakangas, A., Kankkunen, E., Hongisto, V., Virjonen, P., Oliva, D., Keskinen, E., 2011. Effects of five speech masking sounds on performance and acoustic satisfaction. Implications for open-plan offices. Acta Acust. Acust. 97 (4), 641–655. https://doi.org/10.3813/AAA.918444.

Haapakangas, A., Hongisto, V., Hyönä, J., Kokko, J., Keränen, J., 2014. Effects of unattended speech on performance and subjective distraction: the role of acoustic design in open-plan offices. Appl. Acoust. 86, 1–16. https://doi.org/10.1016/j.apacoust.2014.04.018.

Haapakangas, A., Hongisto, V., Liebl, A., 2020. The relation between the intelligibility of speech and cognitive performance - a revised model. Indoor Air 1–17. https://doi.org/10.1111/ina.12726.

Haka, M., Haapakangas, A., Keränen, J., Hakala, J., Keskinen, E., Hongisto, V., 2009. Performance effects and subjective disturbance of speech in acoustically different office types - a laboratory experiment. Indoor Air 19 (6), 454–467. https://doi.org/10.1111/j.1600-0668.2009.00608.x.

Hellbrück, J., Kuwano, S., Namba, S., 1996. Irrelevant background speech and human performance: is there long-term habituation? Journal of the Acoustical Society of Japan (E) 17 (5), 239–247. https://doi.org/10.1250/ast.17.239.

Hockey, G.R.J., 1997. Compensatory control in the regulation of human performance under stress and high workload: a cognitive-energetical framework. Biol. Psychol. 45, 73–93. https://doi.org/10.1016/S0301-0511(96)05223-4.

Hongisto, V., 2005. A model predicting the effect of speech of varying intelligibility on work performance. Indoor Air 15 (6), 458–468. https://doi.org/10.1111/j.1600-0668.2005.00391.x.

Hughes, R.W., Vachon, F., Jones, D.M., 2007. Disruption of short-term memory by changing and deviant sounds: support for a duplex-mechanism account of auditory distraction. J. Exp. Psychol. Learn. Mem. Cogn. 33 (6), 1050–1061. https://doi.org/10.1037/0278-7393.33.6.1050.

Hughes, R.W., Hurlstone, M.J., Marsh, J.E., Jones, D.M., Vachon, F., 2013. Cognitive control of auditory distraction: impact of task difficulty, foreknowledge, and working memory capacity supports duplex-mechanism account. J. Exp. Psychol. Hum. Percept. Perform. 39 (2) https://doi.org/10.1037/a0029064.

ISO, 2003a. ISO 226:2003 Acoustics — Normal Equal-loudness-level Contours.

ISO, 2003b. ISO/TS 15666 Acoustics - Assessment of Noise Annoyance by Means of Social and Socio-Acoustic Surveys. International Organization of Standardization.

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

Jansson, T., 2015. Muumipeikko ja pyrstötähti - Audiobook Spoken by Mr. Panu Kangas. Werner Söderström Ltd.

Jones, D.M., Macken, W.J., 1993. Irrelevant tones produce an irrelevant speech effect: implications for phonological coding in working memory. J. Exp. Psychol. Learn. Mem. Cogn. 19 (2), 369–381. https://doi.org/10.1037/0278-7393.19.2.369.

Kim, H.G., Cheon, E.J., Bai, D.S., Lee, Y.H., Koo, B.H., 2018. Stress and heart rate variability: a meta-analysis and review of the literature. Psychiatry Investig. 15 (3), 235–245. https://doi.org/10.30773/pi.2017.08.17.

Körner, U., Röer, J.P., Buchner, A., Bell, R., 2017. Working memory capacity is equally unrelated to auditory distraction by changing-state and deviant sounds. J. Mem. Lang. 96, 122–137. https://doi.org/10.1016/j.jml.2017.05.005.

Kristiansen, J., Mathiesen, L., Nielsen, P.K., Hansen, Å.M., Shibuya, H., Petersen, H.M., Lund, S.P., Skotte, J., Jørgensen, M.B., Søgaard, K., 2009. Stress reactions to cognitively demanding tasks and open-plan office noise. Int. Arch. Occup. Environ. Health 82 (5), 631–641. https://doi.org/10.1007/s00420-008-0367-4.

Laine, M., Virtanen, P., 1999. Word Mill Lexical Search Program. University of Turku, Centre for Cognitive Neuro Science.

Lee, G.-S.S., Chen, M.-L.L., Wang, G.-Y.Y., 2010. Evoked response of heart rate variability using short-duration white noise. Autonomic Neuroscience: Basic & Clinical 155 (1–2), 94–97. https://doi.org/10.1016/j.autneu.2009.12.008.

Liebl, A., Assfalg, A., Schlittmeier, S.J., 2016. The effects of speech intelligibility and temporal-spectral variability on performance and annoyance ratings. Appl. Acoust. 110, 170–175. https://doi.org/10.1016/j.apacoust.2016.03.019.

Macken, W., Tremblay, S., Alford, D., Jones, D., Macken, B., 1999. Attentional selectivity in short-term memory: similarity of process, not similarity of content, determines disruption. Int. J. Psychol. 34 (6) https://doi.org/10.1080/002075999399639.

Marois, A., Marsh, J.E., Vachon, F., 2019. Is auditory distraction by changing-state and deviant sounds underpinned by the same mechanism? Evidence from pupillometry. Biol. Psychol. 141, 64–74. https://doi.org/10.1016/j.biopsycho.2019.01.002.

Marsh, J.E., Ljung, R., Jahncke, H., MacCutcheon, D., Pausch, F., Ball, L.J., Vachon, F., 2018. Why are background telephone conversations distracting? J. Exp. Psychol. Appl. 24 (2), 222–235. https://doi.org/10.1037/xap0000170.

Muhammad, I., Vorländer, M., Schlittmeier, S.J., 2019. Audio-video virtual reality environments in building acoustics: an exemplary study reproducing performance results and subjective ratings of a laboratory listening experiment. J. Acoust. Soc. Am. 146 (3) https://doi.org/10.1121/1.5126598.

Owen, A.M., McMillan, K.M., Laird, A.R., Bullmore, E., 2005. N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. Hum. Brain Mapp. 25 (1), 46–59. https://doi.org/10.1002/hbm.20131.

Radun, J., Maula, H., Rajala, V., Scheinin, M., Hongisto, V., 2021. Speech is special. The stress effects of speech, noise, and silence during tasks requiring concentration. Indoor Air 31, 264–274. https://doi.org/10.1111/ina.12733.

Röer, J.P., Bell, R., Buchner, A., 2014. Evidence for habituation of the irrelevant-sound effect on serial recall. Mem. Cogn. 42 (4), 609–621. https://doi.org/10.3758/s13421-013-0381-y.

Salamé, P., Baddeley, A., 1987. Noise, unattended speech and short-term memory. Ergonomics 30 (8), 1185–1194. https://doi.org/10.1080/00140138708966007.

Schlittmeier, S.J., Hellbrück, J., Thaden, R., Vorländer, M., 2008. The impact of background speech varying in intelligibility: effects on cognitive performance and perceived disturbance. Ergonomics 51 (5), 719–736. https://doi.org/10.1080/00140130701745925.

Schlittmeier, S.J., Weißgerber, T., Kerber, S., Fastl, H., Hellbrück, J., 2012. Algorithmic modeling of the irrelevant sound effect (ISE) by the hearing sensation fluctuation strength. Atten. Percept. Psychophysiol. 74, 194–203. https://doi.org/10.3758/s13414-011-0230-7.

Sim, C.S., Sung, J.H., Cheon, S.H., Lee, J.M., Lee, J.W., Lee, J., 2015. The effects of different noise types on heart rate variability in men. Yonsei Med. J. 56 (1), 235–243. https://doi.org/10.3349/ymj.2015.56.1.235.

Sörqvist, P., 2010. High working memory capacity attenuates the deviation effect but not the changing-state effect: further support for the duplex-mechanism account of auditory distraction. Mem. Cogn. 38 (5), 651–658. https://doi.org/10.3758/MC.38.5.651.

Szalma, J.L., Hancock, P.A., 2011. Noise effects on human performance: a meta-analytic synthesis. Psychol. Bull. 137 (4), 682–707. https://doi.org/10.1037/a0023987.

Tafalla, R.J., Evans, G.W., 1997. Noise, physiology, and human performance: the potential role of effort. J. Occup. Health Psychol. 2 (2), 148–155. https://doi.org/10.1037/1076-8998.2.2.148.

Tarvainen, M.P., Niskanen, J.P., Lipponen, J.A., Ranta-aho, P.O., Karjalainen, P.A., 2014. Kubios HRV - heart rate variability analysis software. Comput. Methods Prog. Biomed. 113, 210–220. https://doi.org/10.1016/j.cmpb.2013.07.024.

Thayer, J.F., Yamamoto, S.S., Brosschot, J.F., 2010. The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors. Int. J. Cardiol. 141 (2), 122–131. https://doi.org/10.1016/j.ijcard.2009.09.543.

Turner, M.L., Engle, R.W., 1989. Is working memory capacity task dependent? J. Mem. Lang. 28, 127–154. https://doi.org/10.1016/0749-596X(89)90040-5.

Voss, A., Schroeder, R., Heitmann, A., Peters, A., Perz, S., 2015. Short-term heart rate variability - influence of gender and age in healthy subjects. PLoS One 10 (3). https://doi.org/10.1371/journal.pone.0118308.

Weinstein, N.D., 1978. Individual differences in reactions to noise: a longitudinal study in a college dormitory. J. Appl. Psychol. 63 (4), 458–466. https://doi.org/10.1037/0021-9010.63.4.458.

Yadav, M., Cabrera, D., Kim, J., Fels, J., de Dear, R., 2021. Sound in occupied open-plan offices: objective metrics with a review of historical perspectives. Appl. Acoust. 177, 107943 https://doi.org/10.1016/j.apacoust.2021.107943.