COGNITIVE LOAD UNDER URBAN SOUNDSCAPES: A PHYSIOLOGICAL APPROACH USING DIGIT SPAN TASKS

Margret Sibylle Engel[1], Bin Gao[1]

[1]*Chair of Acoustics and Haptics, Dresden University of Technology, Germany*

Abstract

The effect of soundscape on cognitive performance remains an active area of inquiry in psychoacoustics, yet few studies combine behavioural measures with continuous physiological monitoring. This study is the first part of an audiovisual experiment which examines whether urban soundscape conditions modulate cognitive load during a digit span task, as reflected in performance outcomes and autonomic physiological responses. Twenty-seven participants (university students and academic staff) performed forward and backward digit span tasks while exposed to four of five urban soundscape conditions, including traffic, traffic masked with water sounds, traffic masked with simple music, traffic masked with sleep music, and traffic with dynamic loudness variations. Each participant was exposed to four of the five soundscape conditions. Electrodermal activity and heart rate were recorded continuously throughout. Task periods were preceded by two-minute silent baseline recordings, enabling baseline-corrected physiological comparison between forward and backward task conditions. Forward and backward digit span scores are compared across soundscape conditions as a measure of short-term memory retention and working memory manipulation, respectively. Skin conductance response frequency and mean heart rate, corrected to a silent baseline, are examined as indicators of autonomic arousal during each task condition. The higher cognitive demand of the backward task is expected to produce both lower performance scores and elevated physiological arousal relative to the forward condition. This methodological framework contributes to understanding how urban soundscapes interact with cognitive load, offering a combined behavioural and physiological perspective relevant to soundscape design in work and study environments.

Corresponding author: *margret.engel@mailbox.tu-dresden.de*

Keywords: *Urban soundscape, Audiovisual scenes, Digit Span Task, ECG, EDA.*

1. Introduction

The concept of working memory emerged in cognitive psychology during the 1960s [1], with the current, most widely used model formally established by Baddeley and Hitch in 1974 [2]. Cognitive tests allow researchers and clinicians to assess how attention, short-term memory, and working memory are functioning. One common way to evaluate these functions is the digit span task [3]: the forward version reflects short-term memory capacity and basic attention, while the backward version places greater demand on working memory and executive function, since it requires actively manipulating information during mental processing. Working memory is essential to everyday cognitive activities such as reading comprehension [4], following instructions, performing arithmetic, and planning, making it a foundational skill that underlies both educational and professional performance [5].

Soundscape interventions, which allow deliberate additions or modifications of ambient sound, such as introducing natural elements (birdsong, water), masking unwanted noise, or curating context-appropriate acoustic environments [6, 7] are central to soundscape studies because they shift the focus from mere noise reduction to actively shaping how people perceive, feel, and function within an acoustic environment [8]. Their potential use for memory retrieval is largely theoretical rather than empirically established within the soundscape literature itself: most direct evidence for sound's effect on retrieval comes not from ISO 12913-style soundscape-intervention research, but from the separate background-music and auditory-cognition literature [9, 10], which shows that sound can either support or impair recall depending on emotional content, semantic load, and match to the encoding context. Extrapolating from this adjacent body of work and from classic context-dependent memory theory [11, 12], one could hypothesise that reinstating an encoding-congruent soundscape or reducing stress and cognitive load through calming acoustic design might support retrieval, while distracting or attention-capturing interventions might impair it. However, no study has yet tested a soundscape *intervention*, specifically defined and evaluated under the ISO 12913 framework, at the point of memory retrieval; this represents an identifiable gap rather than a confirmed effect. One that a study designed around this precise intersection would be positioned to address directly, with applications in education, clinical memory support, and forensic recall contexts.

By jointly assessing cognitive and physiological measures, this study aims to provide empirical evidence on the extent to which urban soundscapes influence cognitive functioning and to examine which physiological parameters show measurable responses to different soundscape conditions. These findings contribute to the broader soundscape literature on the relationships between acoustic environmental conditions and human cognitive and physiological states.

2. Theoretical Background

2.1 Heart rate activity and electrocardiogram recordings

Heart rate variability (HRV) refers to the natural variation in the time intervals between successive heartbeats, reflecting the dynamic interplay between the sympathetic and parasympathetic branches of the autonomic nervous system as they regulate cardiac activity; it is widely used as a non-invasive marker of autonomic flexibility, stress reactivity, and overall physiological regulation [13]. Among the standard time-domain HRV metrics, SDNN (Standard Deviation of NN intervals) quantifies the overall variability of all normal-to-normal beat intervals over a given recording period, capturing the combined influence of both sympathetic and parasympathetic activity, and is considered a global index of autonomic regulation particularly informative over longer recording durations [13]. RMSSD (Root Mean Square of Successive Differences), by contrast, measures the beat-to-beat variability between adjacent NN intervals and is considered a more specific index of parasympathetic (vagal) activity, making it especially sensitive to short-term changes in relaxation, recovery, or acute stress responses [13]. Together, SDNN and RMSSD are commonly used in psychophysiological research, including studies of environmental sound and stress.

2.2 Nervous system subdivision and electrodermal activity

The nervous system is divided into two main components: the central nervous system (CNS) and the peripheral nervous system (PNS). The CNS consists of the brain and spinal cord, which serve as the primary centres for processing and integrating information. The PNS comprises the network of nerves that connects the CNS to the rest of the body. One of its subdivisions is the autonomic nervous system (ANS), which regulates involuntary physiological functions such as heart rate, respiration, and digestion.

The ANS is further divided into sympathetic and parasympathetic nervous systems. The sympathetic nervous system mediates the fight-or-flight response, increasing heart rate, dilating the pupils, and stimulating sweat gland activity in preparation for action. In contrast, the parasympathetic nervous system promotes the rest-and-digest response by reducing heart rate, facilitating digestion, and supporting recovery and energy conservation. In addition to these structural divisions, the nervous system comprises neurons, specialised cells that transmit electrical impulses and chemical signals across synapses [14].

Electrodermal activity (EDA) measures subtle changes in the skin's electrical conductance that result from sympathetic nervous system activity. Because sweat glands are innervated exclusively by sympathetic fibres, changes in sweat secretion alter the skin's electrical conductance. The fingers, palms, wrists, soles of the feet, and forehead are the most common sites for EDA measurement due to their high density of eccrine sweat glands [15].

EDA consists of two primary components: the skin conductance level (SCL) and the skin conductance response (SCR). SCL represents the slow-varying, tonic baseline level of skin conductance and reflects an individual's general level of physiological arousal. In contrast, SCR consists of rapid, transient (phasic) fluctuations in skin conductance that occur in response to specific internal or external stimuli [15]. Accordingly, EDA can be expressed as:

$$\text{EDA} = \text{SCL (tonic baseline)} + \text{SCRs (phasic responses)} \qquad (1)$$

The common parameters to characterise the tonic baseline component are: 1) SCL is the tonic, background component of skin conductance. It represents a slow-drifting baseline that shifts gradually across tens of seconds to minutes rather than in quick bursts. It tracks your overall level of autonomic arousal, alertness, or sustained emotional engagement, so a rising SCL points to heightened activation while a lower SCL suggests a calmer, more resting state. It's usually extracted from the raw EDA signal by applying a low-pass filter, typically below 0.1 Hz, which smooths away the quicker, spike-like fluctuations and leaves the slow underlying trend. 2) SCL slope, which captures how fast the baseline conductance is changing over time, showing whether the background arousal is rising, falling, or holding steady. A positive slope means baseline arousal or stress is building over that stretch of time, a negative slope means the body is winding down toward a more relaxed state, and a slope near zero means arousal is staying stable. 3) SCL window is simply the chunk of time over which the SCL value or its slope gets calculated. Since SCL moves slowly compared to the sharp, brief spikes of phasic SCRs, working in windows lets researchers follow arousal trends without those quick blips throwing off the picture.

The phasic components are characterised by the following parameters: 1) SCR Amplitude/Peak Amplitude, which indicates how large the skin conductance jump is, in microSiemens, measured from the low point just before the response to its highest point. It reflects the strength of the physiological reaction and is commonly used as an index of arousal or of how salient a given stimulus was to the person. 2) SCR Frequency represents how many skin conductance responses occur within a given time window, whether they're tied to a specific event or arise spontaneously during rest with no clear trigger. It's often used as a marker of underlying stress, mental effort, or general activation of the nervous system. 3) SCR Onset is the moment skin conductance starts climbing after a stimulus is presented. It marks the reaction's latency and typically shows up somewhere between 1 and 3 seconds after the triggering event. 4) SCR Risetime is the span between when the response begins and when it hits its highest point. It serves as a gauge of how quickly physiological arousal builds. 5) SCR Prominence is a metric capturing how much a given SCR rises above the surrounding signal, accounting for shifts in the baseline. Peak-detection algorithms rely on it to tell genuine responses apart from random fluctuations in the noise [15].

3. Methods

3.1 Study areas

The audiovisual recordings were conducted in Dresden, the capital of the German state of Saxony, which has a population of approximately 560,000. Four recording locations were selected: P1 and P2 were situated along the Weißeritz River, a tributary of the Elbe River; P2 was recorded at the Hezereiter bridge, near the Hydropower plant Plauenscher Grund, where the water stream was stronger; while P3 was located on the Elbe River. Traffic recordings were collected at P4, a secondary road running along the Elbe riverbank. The locations of all recording sites are shown in Figure 1.

Figure 1. Audiovisual recording spots

3.2 Methodological procedures

The methodological procedure comprised seven steps, as illustrated in Figure 2.

1) Fieldwork. Measurement locations were selected during the fieldwork phase (Figure 1). Audiovisual recordings were performed using a Squadriga III (HEAD acoustics GmbH), an AMBEO VR microphone (Sennheiser), a Zoom H6 Essential recorder, and an Insta360 Pro camera. All devices were positioned at a height of 1.2 m to simulate the seated position of a participant [16].

2) Audiovisual scene preparation. The recorded audio was edited in Audacity to create 60-second soundscapes, including channel editing and mixing required for traffic-noise masking. Three masking conditions were created using: (i) stream water sounds recorded at P2, (ii) Music Sleep (Max Richter – Dream 1: Before the Wind Blows It All Away, Pt. 8), and (iii) Music Simple (Yutaka Hirasaka – Simple Words).To increase masking effectiveness, the water sounds were amplified by 6 dB, while both music tracks were amplified by 10 dB. An additional condition (iv), referred to as Psychoacoustic, was created by applying a 10 dB adjustment following the procedure described by Sethi et al. (2024) [17].

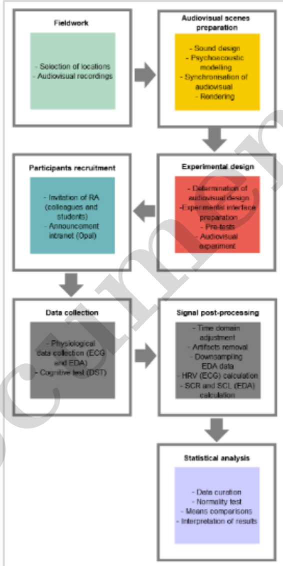

Figure 2. Methodological procedures

After audio editing, all soundtracks were synchronised with the traffic-flow video recorded at P4 and rendered using the open-source software Kdenlive [16]. The psychoacoustic characteristics of each audio stimulus were quantified using the single-value psychoacoustic metrics available in Artemis Suite® (HEAD acoustics GmbH) [16].

3) Experimental design. Each experimental session began with an explanation of the study, completion of the informed consent form, participant pseudo-anonymisation, collection of demographic information, and a training session to familiarise participants with the cognitive tests and the autobiographical memory questionnaire. The autobiographical memory data – soundscape assessment (highlighted in red in Figure 3) are presented separately in the study entitled "When Sounds Bring Back Memories: A Psychophysiological Approach".

As shown in Figure 3, four experimental sessions were conducted to collect cognitive, physiological, and soundscape assessment data. In every session, participants viewed the same video while listening to different audio conditions.

The order of the audiovisual scenes was randomised to minimise expectation effects; therefore, the sequence shown in Figure 6 is illustrative only. A 15-second break between sessions allowed the experimenter to load the next audiovisual scene and update the cognitive testing interface.

Each session consisted of a 2-minute baseline recording without audiovisual stimulation. The tasks and soundscape assessment had audiovisual stimulation. The first task was a forward (FW) Digit Span Task (DST) lasting up to 2 minutes. A second 2-minute baseline preceded the backward (BW) DST, followed by a third 2-minute baseline and the soundscape assessment, which typically lasted approximately 5 minutes. After completing all four sessions, participants were invited to provide feedback on the experiment. The total duration of each experiment was approximately 70 minutes.

The experimental interface was developed in MATLAB R2024b using a graphical user interface (GUI). Prior to data collection, pilot tests and calibration of the audiovisual reproduction system were conducted using an RME Fireface UCX sound card, Sennheiser HD 650 headphones, and a dummy head.

4) Participant recruitment. Participants were recruited by inviting students and research assistants from the Chair of Acoustics and Haptics at TU Dresden and through announcements posted on the TU Dresden OPAL intranet platform.

5) Data collection. Physiological recordings were acquired simultaneously during the baseline periods, Digit Span Tasks, and soundscape assessments. Electrodermal activity (EDA) and electrocardiogram (ECG) signals were recorded using a Biosignalsplux system equipped with a four-channel hub. Signals were sampled at 500 Hz with 16-bit resolution. Ag-AgCl electrodes were placed on the wrist for EDA measurements and at the end of the subject's ribs for ECG measurements. Simultaneously, the MATLAB-based experimental interface automatically recorded participants' responses during the cognitive tasks and soundscape assessment. The outputs were saved in separate Excel files for each task and included the number of digits presented, the number of correct responses, and the DST response times.

6) Signal post-processing. Physiological data were post-processed before analysis. Because of the recording system configuration, time-domain adjustments were required to convert recorded observations into time-referenced signals. Each physiological channel was visually inspected to identify motion artefacts or recording errors that could bias the calculated averages. Detected artefacts were removed prior to further analysis.

For EDA processing, an anti-aliasing low-pass filter was applied before downsampling the signal from 500 Hz to 32 Hz. These preprocessing steps ensured reliable estimation of electrodermal parameters. Subsequently, heart rate variability (HRV) with SDNN and RMSSD; skin conductance response (SCR) with Delta SCR onset, risetime, peak amplitude, amplitude, peak prominence and frequency; and Delta skin conductance level (SCL) slope and window metrics were calculated. By definition, the delta (Δ) is the difference between the task and baseline.

7) Statistical analysis. Finally, statistical analyses were performed on datasets curated according to each research hypothesis. The Shapiro–Wilk test was used to assess the normality of each variable and determine the appropriate statistical test for comparing groups. Because the five audiovisual scenes were randomly distributed across four experimental sessions, the number of observations differed between conditions. Given the uneven sample sizes across conditions, Skillings-Mack tests were applied, as this nonparametric method accommodates missing or unbalanced data [18].

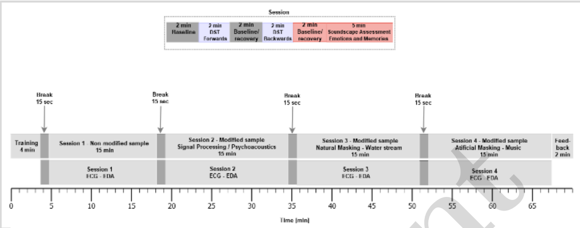

Figure 3. Experimental design

4. Results

4.1 Demographics

The participants comprised 27 persons, academic staff and students, most of them male (74.07%, 20), with an age range of 19 to 63 years, M = 29.85 years, and SD = 8.79 years. Most participants (59.26%) were between 20 and 29 years old. The participants were originally from three continents: 1) Asia (59.6%, 16), 2) Europe (33.3%, 9), 3) South America (3.70%, 1). Data from two participants were excluded from the sample due to technical issues and personal reasons.

4.2 Psychoacoustic characteristics of audio samples

Table 1 shows the average values of the psychoacoustic indicators and Loudness N5 percentile for each presented scene.

Table 1. Psychoacoustic single values of each audio sample

Scene	NAvg soneGF	N_5 soneGF	S acum	R asper	FS vacil	T tuHM
Traffic	38.63	52.18	0.91	0.408	0.014	2.005
Water	36.82	39.90	1.52	0.085	0.006	0.037
Music Simple	10.33	15.30	1.34	0.067	0.004	0.054
Music Sleep	36.18	44.84	1.09	0.247	0.009	-
Psychoac.	43.38	53.84	1.05	0.343	0.092	1.944

The Psychoacoustic condition exhibited the highest loudness and fluctuation strength, whereas the Water condition showed the greatest sharpness. The Traffic condition was characterised by higher tonality and roughness, while the Music Simple condition presented the lowest loudness.

4.3 Digit Span Task Results

As shown in Table 2, forward digit span accuracy was highest under the Music-Sleep and Psychoacoustic conditions, both exceeding performance in the Traffic condition. For the backward digit span task, only the Music-Simple condition outperformed Traffic.

Table 2. Digit Span Task Results

Condition	FW n	FW Accuracy	BW n	BW Accuracy	Missing data (technical glitch)
Traffic	23	93.50%	23	94.50%	ID 5: FW & BW
Water	26	93.00%	26	92.50%	none
Music Simple	11	93.10%	10	95.00%	ID 25: BW
Music Sleep	14	96.00%	14	94.20%	none
Psychoac.	24	97.00%	25	93.00%	ID 6: FW

4.4 Comparison of all conditions

Skilling-Mack tests were conducted to examine differences in physiological delta measures (post- minus pre-task) across three acoustic conditions (Traffic, Psychoacoustic, Water) during the forward and backward Digit Span tasks. This non-parametric approach was selected due to unequal numbers of observations per condition. Holm correction was applied to control for multiple comparisons across the 11 physiological variables within each task.

No significant differences were found across conditions for any physiological measure of the Digit Span Task forwards (Table 3) and backwards (Table 4), where all $p > .12$, and all $p\_Holm = 1$.

Table 3. Skilling-Mack Test Results for Physiological Measures for Digit Span Forwards

Variable	χ^2	df	p	p_Holm	Significant
Δ SDNN	1.264	2	0.532	1	No
Δ RMSSD	4.324	2	0.115	1	No
Δ SCR onset	1.168	2	0.558	1	No
Δ SCR risetime	2.398	2	0.301	1	No
Δ SCR peak amplitude	0.331	2	0.847	1	No
Δ SCR peak prominence	0.147	2	0.929	1	No
Δ SCR amplitude	0.615	2	0.735	1	No
Δ SCR frequency	1.456	2	0.483	1	No
ΔSCL	2.032	2	0.362	1	No
Δ SCL slope	1.899	2	0.387	1	No
Δ SCL window	1.863	2	0.394	1	No

Table 4. Skilling-Mack Test Results for Physiological Measures for Digit Span Backwards

Variable	χ^2	df	p	p_Holm	Significant
Δ SDNN	2.036	2	0.361	1	No
Δ RMSSD	2.639	2	0.267	1	No
Δ SCR onset	1.439	2	0.487	1	No
Δ SCR risetime	4.283	2	0.117	1	No
Δ SCR peak amplitude	0.322	2	0.851	1	No
Δ SCR peak prominence	1.971	2	0.373	1	No
Δ SCR amplitude	3.001	2	0.223	1	No
Δ SCR frequency	0.555	2	0.757	1	No
Δ SCL	0.965	2	0.617	1	No
Δ SCL slope	1.458	2	0.482	1	No
Δ SCL window	0.01	2	0.995	1	No

These findings indicate that the type of background acoustic condition did not differentially affect cardiac (SDNN, RMSSD) or electrodermal (SCR, SCL) reactivity during either digit span task.

5. Discussion

The present study examined whether physiological indices of autonomic reactivity, heart rate variability (SDNN, RMSSD) and electrodermal (SCR and SCL parameters), differed as a function of acoustic background condition (traffic, psychoacoustic, water) during forward and backward Digit Span performance. Contrary to expectations, Skilling-Mack tests revealed no statistically significant differences across conditions for any physiological measure, in either task version, even prior to Holm correction. These null findings warrant careful interpretation.

One important methodological consideration relates to the acoustic properties of the stimuli themselves. It is possible that the Water sound stimulus presented similar low-frequency spectral patterns to the traffic noise condition. If the two stimuli overlapped substantially in their low-frequency energy content, participants may have perceptually or physiologically processed the Water sound in a manner similar to Traffic noise, rather than as a distinct, calming acoustic stimulus.

As a limitation, the use of the Skilling-Mack test was necessitated by unequal sample sizes/missing data across conditions, which can reduce power relative to fully balanced repeated-measures designs. With only 2 degrees of freedom and moderate sample sizes typical of psychophysiological studies, the tests may have been underpowered to detect small-to-moderate effects, particularly for noisy measures like SCR parameters.

6. Conclusions

The aim of this study was to identify which EDA and ECG parameters respond reliably to soundscape-based memory retrieval under cognitive load, with the broader goal of understanding whether autonomic and electrodermal measures could serve as sensitive markers of this process. Overall, the present findings do not support the hypothesis that Traffic, Psychoacoustic, and Water sound conditions differentially affect cardiac or electrodermal. Several limitations faced throughout the experiment likely constrained this observation, including a modest sample size, technical glitches during data collection, and an uneven number of samples across conditions, all of which reduce statistical power and may have obscured genuine but subtle physiological responses.

These preliminary results should not be read as evidence that soundscape interventions have no effect, but rather as a signal that larger, more carefully controlled studies are needed to properly capture what may be subtle physiological effects. Future work should aim for a larger and more balanced sample across conditions to increase statistical power and should consider testing water sounds with varying stream flow rates, since the water stimulus used here may have shared acoustic properties with the traffic condition that made the two harder to distinguish physiologically. Testing additional EDA electrode placements would also help determine whether the effects observed are consistent across recording sites or specific to the one used in this study. Continuing this line of research is worthwhile, as understanding whether and how soundscape interventions influence autonomic and electrodermal activity, particularly in the context of memory retrieval and cognitive load, could have meaningful implications for designing acoustic environments that support relaxation, attention, or wellbeing in the populations being studied.

References

[1] R. C. Atkinson and R. M. Shiffrin, "*Human memory: A proposed system and its control processes.*" In Psychology of learning and motivation (Vol. 2, pp. 89-195). Academic press, 1968.

[2] A. Baddeley and G. Hitch, "Working memory". *Psychol. Learn. Motiv.* 8 47–89, 1974. 10.1016/j.cub.2009.12.014

[3] D. Wechsler, "*Wechsler Adult Intelligence Scale*" (4th ed.). Pearson, 2008.

[4] M. Daneman and P. A. Carpenter, "Individual differences in working memory and reading". *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450–466, 1980.

[5] F. Sana and B. Fenesi, "Working Memory and Instructional Fit: Reintroducing Aptitude–Treatment Interaction in Education Research". *Behavioral Sciences*, 15(6), 765, 2025.

[6] ISO 12913-1:2014 / ISO/TS 12913-4 (in development) — *Acoustics — Soundscape — Part 1: Definition and conceptual framework; Part 4: Design and intervention.* International Organization for Standardization.

[7] T. Van RenterghemT., K. Vanhecke, K. Filipan, K. Sun, T. De Pessemier, B. De Coensel, ... & D. Botteldooren, "Interactive soundscape augmentation by natural sounds in a noise polluted urban park". *Landscape and urban planning*, 194, 103705, 2020.

[8] J. Kang, "Soundscape in city and built environment: current developments and design potentials". *City and Built Environment*, 1(1), 1, 2023.

[9] L. Jäncke, "The effect of background music on episodic memory and autonomic responses: listening to emotionally touching music enhances facial memory capacity". *Scientific Reports,* 2015.

[10] C. A. Ermert, M. Yadav, J. E. Marsh, S. J. Schlittmeier, T. W. Kuhlen, and J. Fels, "Serial recall in spatial acoustic environments: irrelevant sound effect and spatial source alternations". *Scientific Reports*, 15(1), 32473, 2025.

[11] D. R. Godden and A. D. Baddeley, "Context-dependent memory in two natural environments: On land and underwater". *British Journal of Psychology, 66*(3), 325–331, 1975.

[12] E. Tulving and D. M. Thomson, "Encoding specificity and retrieval processes in episodic memory". *Psychological Review, 80*(5), 352–373, 1973.

[13] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms". *Frontiers in Public Health, 5*, 258, 2017.

[14] D. Purves, G. J. Augustine, D. Fitzpatrick, W. C. Hall, A.-S. LaMantia, R. D. Mooney, M. L. Platt and L. E. White (Eds.), *Neuroscience* (6th ed.). Oxford University Press / Sinauer Associates, 2018.

[15] W. Boucsein: *Electrodermal Activity* (2nd ed.). *Springer Science & Business Media*, 2012. https://doi.org/10.1007/978-1-4614-1126-0

[16] M.S. Engel, M. Fairhurst, and E. Altinsoy, "Verification of appropriate neurophysiological tests to assess auditory-cognitive responses during memory retrieval while reproducing urban soundscape scenes". *In Proceedings of Internoise* (Vol. 272, No. 3, pp. 1723-1734), 2025, October.

[17] A. Sethi, M.S. Engel and E. Altinsoy "Perceptual response analysis based on psychoacoustical changes on acoustic scenes of urban soundscapes". *In Proceedings of DAGA 2024*, 2024.

[18] M. Chatfield, M. and A. Mander, "The Skillings–Mack test (Friedman test when there are missing data)". *The Stata Journal*,9(2), 299-305, 2009.