On the Processing and Perception of Operating Room Soundscapes in Workplace Simulations using Mobile Electroencephalography

von der Fakultät VI - Medizin und Gesundheitswissenschaften
der Carl von Ossietzky Universität Oldenburg
zur Erlangung des Grades und Titels eines
Doktor der Naturwissenschaften (Dr. rer. nat.)

eingereichte Dissertation
von Herrn Marc Rosenkranz
geboren am 10. Juni 1994 in Hamburg

Erstbetreuer und Erstgutachter:
Dr. Martin G. Bleichner

Zweitgutachter:
Prof. Dr. Christoph Herrmann

Zusätzliches Mitglied der Prüfungskommission der Disputation:
Prof. Dr. Stefan Debener

Datum der Disputation:
22.04.2025

Abstract

The operating room (OR) is a high-stake workplace, where the personnel must efficiently operate under time-pressure while being exposed to many distractors. The auditory environment of the OR consist of a complex soundscape that contributes to the cognitive strain experienced by surgical personnel. Capturing how the soundscape is perceived and processed can contribute to our understanding of its detrimental effects. Therefore, this thesis explores how the brain's response to realistic soundscapes, as measured by electroencephalography (EEG), can be utilized to understand perceptual and cognitive processes in complex environments.

In Study I, we aimed to investigate the processing of complex soundscapes during the performance of a concurrent task. Therefore, we computed EEG responses to both continuous soundscapes and transient auditory stimuli, such as alarms, while participants performed a visual-motor task. We demonstrated that event-related potentials (ERPs) to discrete sounds could be reliably measured, providing a robust validation of it's application during dynamic tasks. Additionally, temporal response functions (TRFs) were computed for a continuous OR playback, marking an important advancement in assessing neural responses to naturalistic soundscapes. Notably, the study established that these neural responses could be measured while participants were engaged in a concurrent task.

In Study II, we extended these findings by examining how cognitive demand influences EEG responses, subjective workload, and surgical performance during a simulated surgical task set within a naturalistic OR soundscape. Specifically, the study explored whether ERPs and TRFs varied with demand. To vary demand, participants were asked to remember either two or eight letters before the surgical task, silently repeat them during the surgical task, and retrieve them at the end of the surgical task. To increase the distracting potential of the irrelevant soundscape, we presented spoken letters during the surgical task. Subjective measures indicated that participants felt more distracted during the high-demand condition, yet this increase in perceived distraction was neither reflected in task performance nor in the ERPs and TRFs. However, ERPs to spoken letters exhibited temporal variations, suggesting a potential adaptation effect over the course of the experiment. These findings highlight that a combination of several measurements is beneficial to assess cognitive and emotional changes in complex environments.

Study III further examined the specific impact of irrelevant speech on auditory processing under varying task demands using surgical tasks of differing difficulty. For this, we presented an OR soudscape in combination with speech. We showed that while irrelevant speech did not impair task performance, it significantly increased perceived workload and distraction. Furthermore, higher task demand suppressed ERPs to irrelevant sounds and TRFs to

irrelevant speech, reflecting an allocation of cognitive resources to the task. These findings underscore the cognitive strain imposed by irrelevant speech, particularly under high task demand, and highlight the potential of neural measures like ERPs and TRFs to study effects of auditory processing in demanding environments.

In this thesis, I combined neurophysiological measures, self-reports, and performance metrics to provide a comprehensive understanding of auditory processing in complex environments. By leveraging EEG measures such as ERPs and TRFs, we demonstrated the feasibility of assessing neural responses to realistic soundscapes from controlled laboratory settings to more realistic environments. Our findings shed light on how OR soundscapes are perceived and processed, and how the resulting demand affects the individual. These studies not only inform strategies for managing auditory distractions in the OR but also highlight the potential of EEG as a tool for investigating the effects of distraction in high-stake environments like the OR. Future research can build on this work to further explore the interplay between auditory processing, workload, and performance in real-world scenarios.

Zusammenfassung

Die Chirurgie ist eine medizinische Disziplin, die sich dadurch auszeichnet, dass komplexe motorische Aufgaben oft unter Zeitdruck durchgeführt werden müssen und Fehler schwerwiegende Konsequenzen für die Patient*innen haben können. Operationen erfordern daher ein hohes Maß an Konzentration, was über einen längeren Zeitraum hinweg anstrengend ist und zu Stress führen kann. Zudem ist das chirurgische Personal einer belastenden Geräuschkulisse im Operationssaal (OP) ausgesetzt. Die Geräuschkulisse im OP zeichnet sich durch eine Vielzahl sich überlappender Geräusche aus, wie kontinuierliches Summen von Ventilatoren, das Piepsen eines EKG-Monitors, Alarmgeräusche und Gespräche. Um die Auswirkungen dieser Geräuschkulisse auf das Personal zu erfassen, wurden in der Vergangenheit vor allem Fragebögen und Verhaltensmaße eingesetzt. Diese Methoden haben jedoch Einschränkungen. Fragebögen können beispielsweise nur zu einem bestimmten Zeitpunkt eingesetzt werden, etwa am Ende einer Operation, wodurch eine direkte Reaktion auf bestimmte Geräusche nicht erfasst werden kann. Verhaltensmaße haben in der Vergangenheit zu sehr unterschiedlichen Ergebnissen geführt, was unter anderem an den unterschiedlichen Paradigmen liegt, die unterschiedliche chirurgische Aufgaben und Verhaltensmaße eingesetzt haben. In dieser Dissertation haben wir daher in drei Studien die Elektroenzephalographie (EEG) als kontinuierliches und objektives Maß hinzugezogen, um die Reaktionen auf bestimmte Aspekte der Geräuschkulisse und deren Zusammenhang mit kognitiven Einflüssen zu ermitteln. Zu diesem Zweck wurden zwei EEG-Analysemethoden angewendet. Die erste Methode basierte auf ereigniskorrelierten Potentialen (EKPs), welche die Reaktionen auf spezifische Töne erfassten, und die zweite auf zeitlichen Antwortfunktionen (eng. temporal response function, kurz TRF), welche die Reaktionen auf die kontinuierliche Geräuschkulisse erfasst.

In Studie I untersuchten wir die EEG-Reaktionen auf kontinuierliche Geräuschkulissen und vorübergehende diskrete Töne, wie z. B. Alarme, während die Teilnehmenden eine visuell-motorische Aufgabe, das Spiel Tetris, ausführten. Wir konnten erfolgreich zeigen, dass EKPs auf diskrete Töne zuverlässig gemessen werden konnten, was eine robuste Validierung der neurophysiologischen Methodik ermöglichte. Darüber hinaus haben wir TRFs für die kontinuierliche OP-nahe Geräuschkulisse berechnet, was einen wichtigen Fortschritt bei der Auswertung der neuronalen Reaktionen auf natürliche Geräuschkulissen darstellt. Die Studie hat gezeigt, dass diese neuronalen Reaktionen nicht nur in kontrollierten Umgebungen gemessen werden können, sondern auch, während die Teilnehmenden gleichzeitig mit einer Aufgabe beschäftigt waren.

In der zweiten Studie haben wir untersucht, wie unterschiedliche kognitive Anforderungen die EEG-Reaktionen, die subjektive Arbeitsbelastung und die chirurgische Leistung während

einer simulierten chirurgischen Aufgabe mit einer reellen OP-Geräuschkulisse beeinflussen. Um die Anforderungen zu variieren, wurden die Teilnehmenden gebeten, sich vor der Operation entweder zwei oder acht Buchstaben zu merken, diese während der Operation still zu wiederholen und am Ende der Operation wieder abzurufen. Um das Ablenkungspotenzial der irrelevanten Geräuschkulisse zu erhöhen, präsentierten wir während der Operation gesprochene Buchstaben. Subjektive Messungen ergaben, dass sich die Teilnehmenden unter der Bedingung der hohen Anforderung stärker abgelenkt fühlten, doch spiegelte sich dieser Anstieg der wahrgenommenen Ablenkung weder in der chirurgischen Leistung noch in den EKPs und TRFs wider. Die EKPs für gesprochene Buchstaben wiesen jedoch eine zeitliche Veränderung auf, was auf einen möglichen Anpassungseffekt im Verlauf des Experiments hindeutet. Diese Ergebnisse unterstreichen, dass eine Kombination mehrerer Messungen notwendig ist, um kognitive und emotionale Veränderungen zu erfassen, wenn man sich mit komplexen Umgebungen beschäftigt.

In der dritten Studie wurde die Auswirkung von irrelevanter Sprache während chirurgischer Aufgaben mit unterschiedlichem Schwierigkeitsgrad weiter untersucht. Dafür haben wir eine OP-Geräuschkulisse zusammen mit Sprache abgespielt. Wir konnten zeigen, dass irrelevante Sprache zwar die Aufgabenleistung nicht beeinträchtigte, aber die wahrgenommene Arbeitsbelastung und Ablenkung signifikant erhöhte. Darüber hinaus verringert eine höhere Aufgabenanforderung die EKPs auf irrelevante auditorische Reize und die TRFs auf irrelevante Sprache, was auf eine Zuweisung kognitiver Ressourcen für die Aufgabe hindeutet. Diese Ergebnisse unterstreichen das irrelevante Sprache während unterschiedlicher kognitiver Anforderungen unterschiedlich verarbeitet wird, und verdeutlichen das Potenzial neuronaler Maßen wie EKPs und TRFs zur Untersuchung der Auswirkungen auditiver Ablenkung in anspruchsvollen Umgebungen.

In dieser Arbeit habe ich neurophysiologische Messungen, Fragebögen und Verhaltensmaße kombiniert, um ein umfassendes Verständnis der auditorischen Verarbeitung in komplexen Umgebungen zu gewinnen. Durch den Einsatz von EEG-Korrelaten der auditorischen Verarbeitung wie EKPs und TRFs konnten wir zeigen, dass es möglich ist, die neuronalen Reaktionen auf realistische Geräuschkulissen von kontrollierten Laborumgebungen bis hin zu realistischeren Umgebungen zu erfassen. Unsere Ergebnisse geben Aufschluss darüber, wie OP-Geräuschkulissen wahrgenommen und verarbeitet werden. Diese Studien unterstreichen nicht nur das Potenzial des EEG als Instrument zur Untersuchung der Auswirkungen von Ablenkungen in Umgebungen wie dem OP, sondern geben auch Aufschluss über Strategien zur Bewältigung auditiver Ablenkungen im OP. Zukünftige Forschungen können auf dieser Arbeit aufbauen, um das Zusammenspiel zwischen auditiver Verarbeitung, Arbeitsbelastung und Leistung in realen bzw. realistischen Szenarien weiter zu untersuchen.

Danksagung

Zunächst möchte ich mich bei meinem Erstbetreuer Martin bedanken. Du hast mein Promotionsprojekt von Anfang an mit Begeisterung begleitet, meine Datenkrisen ruhig aufgefangen und auch dafür gesorgt, dass ich genügend Freiraum für eigene Ideen hatte. Durch unseren Austausch von der kleinteiligen Signalanalyse bis hin zu philosophischen Fragen hast Du nicht nur meine fachliche Kompetenz gestärkt, sondern auch immer mein persönliches Interesse im Blick gehabt. Ich möchte Dir für Deine Zeit, Dein Engagement und Deine Supervision danken, unter der ich stets einen abwechslungsreichen und spannenden Arbeitsalltag erlebt habe.

Auch der NELI-Gruppe und den NELI-Alumni möchte ich herzlich danken. Silvia, Thorge, Manu, Arnd und Daniel, es hat wirklich Spaß gemacht mit euch zu arbeiten, zu diskutieren, zu spielen und zu feiern. Auch Stefan möchte ich für seinen wissenschaftlichen Scharfsinn und seine Beteiligung an meiner Entwicklung danken. Gleiches gilt für die gesamte AG Neuropsychologie, bei der ich mich für die freundschaftliche und konstruktive Atmosphäre, die vielen Kuchen, Lab Socials und Abende voller Spaß bedanken möchte. Es war großartig, Zeit in einem Team mit so vielen kompetenten und netten Menschen zu verbringen, sowohl innerhalb als auch außerhalb der Arbeit.

Außerdem möchte ich allen weiteren Kolleg*innen im Department, der Open Science Initiative Oldenburg (OSIG) und meiner HiWi Agnes für die Zusammenarbeit danken.

Danke auch an das Team der Allgemein- und Viszeralchirurgie des PIUS-Hospital unter der Leitung von Prof. Weyhe, deren Türen immer offen standen für interdisziplinären Austausch und spannende Perspektivwechsel.

Ein großer Danke geht auch an meine Familie und Freunde für die vielfältige Unterstützung, das interessierte Nachfragen und das Mitfiebern in allen Phasen meines Lebens. Ganz besonders möchte ich mich bei meinen Eltern Antje und Rolf bedanken. Ihr habt mir immer das Vertrauen gegeben, meine Ziele zu verfolgen.

Zu guter letzt geht noch ein riesiger Dank an meine Frau Merle. Du hast mich durch alle Phasen meiner Promotion und auch lange davor begleitet, dabei stets meine Freuden und Sorgen geteilt und warst eine Stütze, auf die ich mich immer verlassen konnte. Ich habe das Privileg sehr genossen, nur ein paar Treppenstufen gehen zu müssen, um in deinem Büro einen Rat, einen Schnack oder einen Granola-Riegel zu bekommen.

Contents

Glossary

dB(A)	A-weighted decibels
ECG	Electrocardiogram
EDA	Electrodermal activity
EEG	Electroencephalography
EMG	Electromyography
ERP	Event-related potential
GFP	Global field power
GLMM	Generalized linear mixed model
ICA	Independent component analysis
ICC	Intraclass correlation coefficient
IMUs	Inertial measurement units
LMM	Linear mixed model
OR	Operating room
RMS	Root-mean-square
SE	Standard error
SNR	Signal-to-noise ratio
TRF	Temporal-response-function
WHO	World Health Organization

1 General introduction

Some professions demand exceptional skill and precision, where an error can have severe and even fatal consequences. Surgery is one such profession, characterized by high levels of concentration to perform challenging tasks under time-pressure. Surgical procedures can last several hours, thereby physically and mentally straining the surgical team, which usually includes two or more surgeons, an anesthetist and several assisting nurses. Additionally, the acoustic environment is highly complex, comprising a multitude of simultaneous sounds generated by surgical instruments, machinery and verbal communication: Ventilation systems produce a soft and continuous humming sound; an electrocardiogram (ECG) produces a rhythmic sound corresponding to the patient's heartbeat; a surgeon may utilize an electric cutter, which produces a feedback sound to indicate that electricity is running through the instrument; the assisting surgeon may request an instrument from the nurse; two team members situated in the back of the room observe the surgical procedure and engage in conversation about their recent activities when, unexpectedly, a telephone begins to ring. Overall, the operating room (OR) is a workplace where both the task and the environment are complex.

Working in the OR is undeniably demanding. Surgery is characterized by long working hours, time pressure and frequent work interruptions, all of which contribute to increased stress and reduced well-being (Bohrer, Koller, Schlitt, Bauer, & on behalf of the German Society of Surgery, 2011; Kern et al., 2019). Additionally, surgical personnel frequently face shift work, which has negative effects on both mental and physical health (Vogel, Braungardt, Meyer, & Schneider, 2012). Organizational stressors, such as hierarchical structures and high administrative workloads, further impact the working atmosphere and job satisfaction (Bohrer et al., 2011). Beyond these, the nature of surgical procedures themselves is highly demanding. Surgeons must rapidly process information from multiple sensory modalities, including vision, audition, and touch, while making critical decisions that require high levels of concentration. Given these circumstances, it is unsurprising that high levels of stress are reported at all stages of surgeons' careers, which can ultimately lead to severe mental disorders such as burnout (DeCaporale-Ryan et al., 2017; Etheridge et al., 2023; Kern et al., 2019; Lebares et al., 2018; Shanafelt et al., 2015). Moreover, prolonged occupational stress is not only a psychological burden but also a physical one. For example, high rates of hypertension have been observed among surgeons, increasing their risk of cardiovascular disease (Marrelli, Gentile, Palmieri, Paduano, & Tatullo, 2014; Rieger, Stoll, Kreuzfeld, Behrens, & Weippert, 2014).

Among the numerous stressors of surgical personnel, the soundscape of the OR represents a significant stressor that has been identified and studied for many years (Mentis, Chellali, Manser, Cao, & Schwaitzberg, 2016; Shapiro & Berland, 1972). As demonstrated in the introductory scene, personnel are exposed to a constant stream of auditory information, some of which is relevant, such as an alarm or case-relevant communication, and some of which is irrelevant, such as tool clattering and case-irrelevant communication. The relevance of a sound to an individual depends upon their role within the surgical team. For example, the feedback sound of a monitoring device may be relevant to the anesthetist but not for the surgeon. The feedback sound of an instrument may be relevant to the surgeon but not to the anesthetist. Case-irrelevant communication may be irrelevant to the surgeon and anesthetist, but may originate from the teaching of students. Thus, an irrelevant sound to the surgeon may be distracting but cannot be easily avoided as it is relevant to someone else. Moreover, medical devices may generate false alarms, which can be a source of irritation for personnel and contribute to an overall increase in noise levels (Edworthy, 2013). The result of such a soundscape is the presence of numerous potentially distracting sounds that the personnel must process in order to remain alert to those that are relevant. Consequently, the issue of distraction caused by the soundscape of the OR has been a topic of ongoing debate in the literature (Healey, Primus, & Koutantji, 2007; Healey, Sevdalis, & Vincent, 2006; Keller et al., 2018; Padmakumar et al., 2017; Persoon, Broos, Witjes, Hendrikx, & Scherpbier, 2011; Sevdalis, Healey, & Vincent, 2007; Sevdalis et al., 2014; Tsiou, Efthymiatos, & Katostaras, 2008; van Harten, Gooszen, Koksma, Niessen, & Abma, 2021).

Recognizing the challenges posed by the OR soundscape, some studies have attempted to implement interventions aimed at mitigating its distracting effects. For instance, Engelmann, Neis, Kirschbaum, Grote, and Ure (2014) introduced an educational program along with a visual feedback system that indicated whether the intensity levels of the soundscape were high. This approach successfully led to a reduction in overall sound intensity levels. Similarly, Leitsmann et al. (2021) tested a communication device that functioned as a noise-canceling headset while still allowing team communication. In both studies, participants reported perceiving the interventions as beneficial. However, these self-reported improvements were not reflected in objective physiological measures, such as heart rate, nor in performance-related outcomes, such as postoperative outcome. These findings highlight the complexity of assessing the impact of the OR soundscape on personnel. Self-reports only record data from a single point in time, which may result in individuals failing to recognize when they were distracted (Marsh, Bell, Röer, & Hodgetts, 2024). Physiological or postoperative outcome measures may not be sensitive enough to detect the effects of the soundscape due to the many confounding factors in such complex settings, such as heterogeneity of patients and surgical expertise (Engelmann et al., 2014). Furthermore, due to the fast pace of surgery, observations may not capture all potentially distracting events (Ayas, Donmez, Kazlovich, Lombardi, & Jain, 2022). To overcome the limitations of observational, subjective and performance measures for studying human behavior in a work context, researchers have turned to the measurement of brain activity at work. This has led to the emergence of the field of *neuroergonomics* (Parasuraman, 2003), which combines the fields of neuroscience and ergonomics. The neuroergonomic approach includes the study of the neural basis of perceptual and cognitive function, such as hearing and attention, in the workplace or in workplace-related situations. The advantage of studying the brain is that it provides an objective measurement of how personnel react to task- and environmental demands. The ideal method to study how surgeons react to the soundscape is one that is continuous, as this provides real-time monitoring of the brain throughout the surgery and identifies how the brain reacted to sounds. Furthermore, the method should not interfere with the procedure by distracting the surgeon.

Electroencephalographie (EEG) has the potential to fulfill these requirements. EEG measures the electric postynaptic potentials generated by pyramidal neurons in the brain at the scalp of a person. The EEG signals are transmitted to an amplifier, which also digitizes the analogue signal. The development of small and wireless amplifiers fitted to the back of the head enhanced the wearer's mobility and facilitated the acquisition of mobile EEG data beyond the confines of the laboratory setting (Debener, Minow, Emkes, Gandras, & de Vos, 2012). Thus, mobile EEG is a promising method for neuroergonomic assessments (Parasuraman, 2003; Wascher et al., 2021). The high temporal precision of EEG enables the collection of responses to specific situations and stimuli within the environment. Thus, EEG has the potential to measure responses to the complex soundscape of the OR.

Collecting EEG data outside traditional laboratory settings presents several challenges. Participants are embedded in dynamic environments that not only introduce artifacts, both physiological (e.g. eye movements) and non-physiological (e.g. cable movement) (Jacobsen, Blum, Witt, & Debener, 2021; Klug & Gramann, 2021), but may also alter extracted neural parameters (Gramann et al., 2021). Additionally, real-world tasks engage multiple cognitive processes, in contrast to the controlled isolation of specific processes (e.g. attention) in laboratory research (Wascher, Heppner, & Hoffmann, 2014). While theories and models of cognition remain essential, their validation requires replication in more realistic environments. A gradual transition from highly controlled experiments to more naturalistic settings, where task and environmental complexity increase while experimental control decreases, can facilitate the application of laboratory findings to real-world brain measurements (Gramann et al., 2021). This thesis addresses these challenges by introducing intermediate steps between traditional lab-based experiments and uncontrolled environments to investigate auditory distraction in the workplace using EEG.

In the following paragraphs, I will first provide an overview about the different research angles that have so far been used to study auditory distraction in the OR. This will provide a framework for the selection of experimental designs and tasks employed in our studies. Second, I will present the theoretical framework of auditory attention and distraction and the relationship between these concepts and distraction in the OR. Third, I will introduce the EEG parameters we utilized to study auditory processing. Finally, I will provide a summary of the principal aspects of the study of the OR environment and the theoretical background that motivated each of our studies.

1.1 Auditory distraction in the operating room

1.1.1 The soundscape

The soundscape of the OR contains many relevant and irrelevant sounds which accumulate to a constant stream of auditory information. Thus, the sounds vary in many ways, such as content, loudness and spectro-temporal profile. To objectively capture and describe the soundscape many studies have investigated a specific acoustic feature, namely sound intensity. This revealed that the average sound intensity level within and across surgical specialties is 55 A-weighted decibels (dB(A)) (Baltin et al., 2020; Engelmann et al., 2014; Gülşen, Aydıngülü, & Arslan, 2021; Healey et al., 2007, 2006; Keller et al., 2018; Kurmann et al., 2011; Tsiou et al., 2008). The sound intensity level varies between surgical specialties. For instance, orthopaedic surgeries exhibit higher average levels than non-orthopaedic surgeries (Giv, Sani, Alizadeh, Valinejadi, & Majdabadi, 2017; Gülşen et al., 2021; Tsiou et al., 2008). Furthermore, sound intensity levels exhibit variation within surgical specialties. For instance, splenectomy procedures demonstrate higher intensity levels than inguinal hernia surgeries among visceral surgical procedures (Baltin et al., 2020). Additionally, within surgical procedures, sound intensity levels fluctuate between the preparation, main and closing phases of surgery (Keller et al., 2018; Tsiou et al., 2008). By studying sound intensity, these studies show the heterogeneity of the OR soundscape, both within and between surgical procedures. This diversity can be attributed to various factors, including the specific tools employed, the volume of communication, and the complexity of the surgical technique.

It has been demonstrated that high sound intensity levels may pose a risk to personnel and patients. While the World Health Organization (WHO) states that sound intensity levels in hospitals should not exceed 35 dB(A), as this may cause stress to patients, there are no specific guidelines for the OR (Berglund, Lindvall, Schwela, & World Health Organization Occupational and Environmental Health Team, 1999). However, peak sound intensity levels caused by sudden events, such as the drop of a metal tray, have been observed to exceed 80 dB(A), thereby increasing the risk of hearing impairment (Baltin et al., 2020; Engelmann et al., 2014; Giv et al., 2017; Gülşen et al., 2021; Healey et al., 2007; Jeyaraman et al., 2024; Tsiou et al., 2008). Additionally, elevated sound intensity has been associated with an increased risk of surgical site infections (Kurmann et al., 2011), although the causal effect of elevated sound intensity on patient risk remains uncertain. It is plausible that higher sound intensity causes a greater incidence of performance errors, however, it is also possible that more difficult procedures, which heighten the probability of surgical site infections, cause higher sound intensity (Kurmann et al., 2011). Alternatively, sound intensity may exert a direct influence on patients' post-operative outcome, for example, noise-canceling headphones for the patient can reduce post-operative pain (You et al., 2024). Overall, sound intensity provides a descriptive overview of the diversity and dynamics of the OR soundscape and serves as an indicator of potential risks to both personnel and patient health. Establishing clear causal relationships between sound intensity and surgical outcomes remains challenging due to the multifaceted nature of surgical procedures and their environments.

1.1.2 Observational studies and subjective reports

Several studies have examined the impact of the OR soundscape on personnel, either through direct observations or self-reports. The number of auditory and non-auditory distractions observed in the OR has been found to range between 0.2 and 3 distracting events per minute (Ayas et al., 2022; Healey et al., 2007, 2006; Persoon et al., 2011; van Harten et al., 2021). This underscores that surgical staff are regularly exposed to potential distractions throughout surgical procedures. Despite efforts to standardize the observation of distracting events (Healey et al., 2006; Sevdalis et al., 2007, 2014), observers may still fail to identify covert distractions or those that do not elicit a visible behavioral response. This makes it challenging to assess the impact of such distractions on an individual's internal state. Despite this limitation, the majority of studies concur that case-irrelevant communications, such as teaching discussions or telephone calls, are among the most disruptive distractions (Ayas et al., 2022; Healey et al., 2007, 2006; Persoon et al., 2011; Sevdalis et al., 2007, 2014; van Harten et al., 2021). However, some research highlights potential benefits of case-irrelevant communication (Ayas et al., 2022) and that it may not be as disruptive as often assumed (Widmer et al., 2018), as it may increase the team moral.

Observations have also shown that the perception of the soundscape interacts with the workload experienced during surgery. Workload is defined as the amount of cognitive resources a person must invest to achieve optimal task performance (Wickens, 2008). It is influenced by both task demands and environmental demands (Hart & Staveland, 1988). Observational data suggest that while some OR personnel are engaged in high-demand phases of their tasks, others may be in low-demand phases at the same time. Those in low-demand phases are more likely to engage in case-irrelevant communication, which can be distracting to others in a high-demand phase (Ayas et al., 2022; Sevdalis et al., 2007; van Harten et al., 2021; Wheelock et al., 2015). Therefore, the level of demand associated with a distraction and the source of the distraction can vary significantly depending on the specifics of the surgical procedure and the demands placed on a given team member. This highlights the personal nature of distraction, as each individual's threshold for distraction may differ based on their specific circumstances and responsibilities (Keller et al., 2018; Tsiou et al., 2008).

Subjective reports are aligned with observational findings, indicating that case-irrelevant conversations are perceived as the most disruptive sounds, followed by the machinery sounds and ventilation systems (Gülşen et al., 2021; Tsiou et al., 2008; Weigl, Antoniadis, Chiapponi, Bruns, & Sevdalis, 2015). Personnel further report that the OR soundscape can reduce attentiveness (Gülşen et al., 2021) and increase perceived workload (Weigl et al., 2015; Wheelock et al., 2015).

In summary, both observational and subjective data suggest that sounds in the OR, particularly case-irrelevant communication, can be highly disruptive. However, the impact of distractions depends on several factors, including task demand, the roles of team members, and the nature of the distraction itself, highlighting the complexity of distractions in the OR.

1.1.3 Experimental studies

To examine the immediate effects of sound distraction on surgeons, several experimental studies have employed surgical simulators to assess both performance and subjective responses. A common distractor used in these studies is the playback of an OR soundscape. Based on the findings from observational studies and subjective reports, it can be expected that the inclusion of irrelevant speech in the playback is likely to significantly increase the distractive potential of the soundscape. I therefore differentiated between studies based on whether or not speech was included in the OR soundscape.

In studies where speech was absent from the playback or not explicitly mentioned in the description, surgical performance was generally unaffected (Bereuter et al., 2024; Moorthy, Munz, Dosis, Bann, & Darzi, 2003; Moorthy, Munz, Undre, & Darzi, 2004), though exceptions exist (Gao et al., 2018). Dichotic listening to two music pieces, which represents a severe auditory distraction, can decrease performance (Conrad et al., 2010, 2012). This suggests that performance can decrease under severely distracting conditions, but not necessarily due to the non-verbal noise of an OR playback alone.

When task-irrelevant speech was explicitly included in the soundscape or was the sole tested stimulus, the results varied. Szafranski, Kahol, Ghaemmaghami, Smith, and Ferrara (2009) found that an OR soundscape combined with the participant's name being called led to more errors compared to a quiet condition. Similarly, Pluyter, Buzink, Rutkowski, and Jakimowicz (2010) observed increased errors when an OR soundscape with conversation was presented alongside a visual distractor. Siu, Suh, Mukherjee, Oleynikov, and Stergiou (2010) reported that exposure to an OR soundscape with conversations, compared to silence, hindered performance improvement. Czerwiec, Vannier, and Courage (2024) found that ambient chatter increased task completion time relative to a silent condition. Han et al. (2022) showed that team performance declined when an OR soundscape with unintelligible chatter was compared to music. Despite these findings, the available studies allow for limited conclusions about the specific effects of irrelevant speech on surgical performance. In most cases, irrelevant speech was not the primary focus of the studies, which examined noise or distraction effects more broadly. As a result, they often lacked a control condition with an OR playback excluding speech or other distractors. Additionally, the speech stimuli varied considerably, from unintelligible chatter to the participant's name being called out. One study even found no effect on performance when speech was included (Suh et al., 2010). The methodological differences, including variations in tasks and performance measures, as well as the unspecified amount or audibility of speech, make it difficult to determine how speech exposure influences the performance.

Studies that involved task-relevant stimuli, where participants had to respond to speech or sound, consistently demonstrated a decline in surgical performance (Feuerbacher, Funk, Spight, Diggs, & Hunter, 2012; Suh, LaGrange, Oleynikov, & Siu, 2016). The extent of this decrease appears to be related to task complexity, with more difficult tasks being more vulnerable to interference (Yang et al., 2017). Thus, dual-tasking affects surgical performance.

Relatively few studies have explored subjective reports alongside performance data. Those indicate that soundscapes including speech lead to increased feelings of irritation and perceived workload (Gao et al., 2018; Pluyter et al., 2010). When speech was not included, subjective workload remained unaffected (Bereuter et al., 2024). This highlights the significant role that speech can play in shaping subjective experiences.

In summary, despite numerous observational studies identifying irrelevant speech as a major distraction in the OR, experimental evidence on its impact on surgical performance remains limited. Existing evidence suggests that soundscapes with speech may impair surgical task performance compared to those without speech. However, methodological differences across studies, including variations in surgical procedures, performance measures, types of distractors, and the absence of control conditions, make it difficult to draw definitive conclusions. Additionally, given that personnel in high-demand phases report greater distraction from the soundscape, it is notable that the relationship between distraction and task demand has not been systematically tested. Together, these factors complicate the interpretation of how the soundscape affects performance.

1.1.4 Physiological measurements

To further quantify the effects of auditory distraction, several studies have employed objective physiological measures beyond surgical task performance. Electromyography (EMG), which assesses muscle tension, has been shown to increase when participants are exposed to OR soundscape playbacks compared to silence, suggesting a heightened physiological stress response to auditory stimuli (Siu et al., 2010). This increase in EMG activity is even more pronounced when participants must actively respond to auditory distractors, indicating that the cognitive demand of processing and reacting to such stimuli further amplifies the stress response (Suh et al., 2016). However, some motor functions, such as force exertion and movement precision, appear to remain unaffected by OR soundscapes (Bereuter et al., 2024). Other physiological indicators of cognitive load, such as pupil dilation, have also been found to increase significantly in response to OR sound playback, suggesting heightened mental effort or arousal (Gao et al., 2018).

While these effects have been demonstrated under controlled laboratory conditions, applying similar measurements in real surgical settings presents methodological challenges.

For example, the use of noise-canceling headphones in actual surgeries did not result in measurable changes in heart rate (Leitsmann et al., 2021), and a noise reduction intervention that successfully lowered overall sound intensity in the OR had no significant effect on physiological stress markers, such as electrodermal activity (EDA) or cortisol levels (Engelmann et al., 2014). The lack of effect may be due to the fact that heart rate and EDA measuring general stress responses, thus the individual contribution of the soundscape to stress is difficult to extract during the performance of a real surgery.

1.1.5 Implications for our studies

The reviewed literature highlights the complexity of studying distraction in real-life settings like the OR and has motivated several aspects of the current thesis. First, the OR soundscape has a very heterogeneous sound profile across and within surgeries with fluctuating sound intensities around 55 dB(A). The relevance, type of sound, and its intensity can influence its potential to distract and should be considered when constructing an experiment. However, most sounds that are perceived as distracting are irrelevant to the procedure. Therefore, we focused especially on the perception of irrelevant soundscapes. Second, the reported studies imply that the perception of the soundscape is related to the surgical task demand. Although this aspect recurs across observations and reports, it has been neglected in most experimental studies. Third, it is beneficial to capture the complexity of the OR environment using a multi-modal approach (e.g., self-reports, behavioral assessments, and physiological measurements). Furthermore, when using a physiological measurement it would be beneficial if it can be directly related to the soundscape when aiming for real-world recordings, which motivated the use of EEG. Thus, we always collected EEG data, performance data and self-reports to understand how OR soundscapes are processed and perceived, interact with task complexity and influence well-being.

1.2 Models of auditory distraction

Auditory distraction is defined as the interference of an irrelevant or unattended sound with a task being performed. Consequently, the study of auditory distraction can be considered a study of attention. Early theories of attention aimed to identify at which step in the stimulus processing hierarchy relevant stimuli are filtered from irrelevant stimuli (Broadbent, 1958; Treisman, 1969). The critical question was why some stimuli are filtered out early in the processing hierarchy, while other stimuli still pass through the filter. Nowadays, attention is understood to have two principal mechanisms, top-down and bottom-up attention (Corbetta & Shulman, 2002). The former is responsible for regulating goal-directed behavior in accordance with the demands of a given task, whereas the latter is driven by external stimuli. For example, a surgeon focusing on the patient represents a top-down process, whereas an alarm sound engaging attentional resources exemplifies a bottom-up process. Consequently, distraction by an irrelevant stimulus is an interplay between bottom-up stimulus properties and top-down behavioral goals. While these two mechanisms interact, theories of distraction differ in their focus on either bottom-up or top-down processes. Below I will outline the most influential theories of (auditory) distraction and how they relate to our study.

1.2.1 Duplex-mechanism account of auditory distraction

The duplex-mechanism account of auditory distraction (Hughes, 2014) describes how irrelevant auditory stimuli capture attention and thereby interfere with a task. Thus, it focuses especially on bottom-up processes of distraction. It proposes two mechanisms of auditory distraction, namely interference-by-process and attentional capture. According to this theory, these two mechanisms involve different aspects of auditory distraction and should therefore be treated as distinct phenomena.

Interference-by-process refers to the phenomenon whereby an irrelevant sound distracts by engaging similar cognitive processes as the task being performed. This phenomenon is most commonly observed in the context of the changing-state effect. This effect describes that a sequence of irrelevant sounds that vary in a characteristic (e.g. content) - here referred to as the changing-state sequence - is more disruptive than a sequence of repetitive sounds, also called a steady-state sequence. This phenomenon is especially prominent during the performance of a serial recall task. In this task, the participant is required to remember and store a list of items (e.g. words, letters, numbers) in their working memory. The working memory is a cognitive system responsible for storing a limited amount of information for a short period of time. The changing-state sequence (e.g. A D G H) creates a constant change in auditory input which requires more cognitive resources to process than a steady-state sequence (e.g. A A A A), resulting in interference with the serially stored items (Hughes, 2014).

Attentional capture describes the process by which a stimulus engages attentional resources, thereby distracting the person from the main task (Hughes, 2014). The stimulus may be either salient (e.g., one's own name) or a deviant sound embedded in a series of repetitive sounds (e.g., A A A C A). Similar to interference-by-process, this effect is prominent during performance of a serial recall task.

Although the duplex-mechanism account has become the standard model to explain auditory distraction, its critics propose a unitary explanation and suggest that interference-by-process and attentional capture have a shared underlying mechanism (Körner, Röer, Buchner, & Bell, 2017; Marsh et al., 2024). Nevertheless, the duplex-mechanism account has made a significant contribution to our understanding of auditory distraction by providing an explanation for how different types of irrelevant sounds interfere with cognitive tasks: Auditory distraction occurs either when the task and distractor engage similar cognitive resources (i.e., interference-by-process) or when the distractor engages attentional resources (i.e., attentional capture).

1.2.2 Irrelevant speech effect

Although the changing-state effect has been observed in sounds that vary in one dimension (e.g. amplitude), the more complex the sounds become, the greater the disruption they cause (Ellermeier & Zimmer, 2014). This is particularly evident in the case of irrelevant speech, which has been shown to exhibit the highest level of disruption (Ellermeier & Zimmer, 2014). As speech comprises a rapid succession of changing phonological elements, it disrupts the serial storage of elements in working memory (Jones, Madden, & Miles, 1992). Thus, speech must not be understood to be distracting, but speech understanding increases its distracting potential (Little, Martin, & Thomson, 2010; Ueda, Nakajima, Kattner, & Ellermeier, 2019). Therefore, natural irrelevant speech is similar to a complex changing-state sequence. However, it can also act as an attentional capture. For example, when ones name is presented in the ignored stream (Holtze, Jaeger, Debener, Adiloğlu, & Mirkovic, 2021; Moray, 1959), or during whispered speech (Kattner, Föcker, Moshona, & Marsh, 2024), speech results in a shift of attention.

1.2.3 Load Theory

In an attempt to resolve the early/late attentional selection debate, Lavie and Tsal (1994) proposed that the amount of perceptual load required to process target stimuli determines the capacities to process irrelevant or distracting stimuli. Perceptual load is determined by the amount of sensory stimuli that must be processed. When perceptual load is high (e.g., spotting your friend in a crowd) distractor processing is low because the task at hand leaves little capacity for such processing. In other words, when load is high, attention is captured early and distractor processing is reduced, while a low load results in the late processing of irrelevant stimuli, as cognitive capacities are available. The theory was expanded to distinguish between perceptual and cognitive load (Lavie, 2005). Cognitive load is determined by the amount of cognitive resource (e.g. working memory) a task requires. The theory proposes that during high cognitive load, distraction is more likely than during low cognitive load. However, there is ongoing debate as to whether high cognitive load results in high or low susceptibility to distraction (Brockhoff, Schindler, Bruchmann, & Straube, 2022). The evidence suggests that the relationship between the target and the distractor influences the extent to which cognitive load affects distractor processing (SanMiguel, Corral, & Escera, 2008). For example, distractor processing is increased during high compared to low cognitive load when target and distractor stimuli belong to the same sensory modality (Lavie, 2005). The opposite is true when target and distractor stimuli belong to different modalities (Sörqvist, Dahlström, Karlsson, & Rönnberg, 2016).

1.2.4 Distraction theories in the real world

In recent years, research has shifted from controlled laboratory studies to naturalistic designs incorporating real-world stimuli, simulations, and beyond-the-lab investigations (Gramann, Ferris, Gwin, & Makeig, 2014; Hamilton & Huth, 2020; Matusz, Dikker, Huth, & Perrodin, 2019; Vigliocco et al., 2024). However, regarding the changing-state effect and irrelevant speech effect, little evidence exists on how sounds and speech interfere with non-verbal, non-memory tasks (Haapakangas, Hongisto, & Liebl, 2020; Szalma & Hancock, 2011). While working memory is involved in many everyday tasks, including those in the OR, isolating its role in real-world performance is difficult due to the engagement of multiple processes (Sörqvist, 2015). Similarly, in the OR, it is challenging to determine whether irrelevant speech distractions result from the changing-state effect or the deviant effect. For example, a person may audibly discuss the upcoming procedure, creating a changing-state sequence that also contains relevant, attention-grabbing information. Distinguishing between interference-by-process and attention capture may therefore be impractical in the OR but could explain why irrelevant speech is frequently reported as a distraction, motivating further investigation in realistic settings.

A key advantage of traditional distraction paradigms is their clear impact on performance, such as increased errors in serial recall tasks. In contrast, such direct effects may be harder to detect in the OR, as indicated by the discrepancy between frequent self-reports of distraction and mixed performance findings in experimental studies. Given this uncertainty, I will use the term auditory distraction with caution, as it implies a measurable cognitive effect. Instead, I will refer to auditory processing and its relationship to performance measures and self-reports to reflect the unknown extent how the OR soundscape affects the individual.

Cognitive and perceptual load have also been tested in many highly controlled laboratory tasks. However, in real-life it is difficult to differentiate between the specific load processes (Murphy, Groeger, & Greene, 2016). For example, laparoscopic surgery may increase perceptual load compared to open surgery because anatomical landmarks are more difficult to identify. At the same time, laparoscopic surgery may increase cognitive load as it requires navigation of instruments in a 3D space, while the surgeon receives visual input from a 2D screen. This also complicates the prediction from the load theory in the real world, whether load increases or decreases distractor processing. Yet, the load theory provides an important basis for the study of distraction in real-world situations, namely that distractor processing depends on the amount of limited resources available to process a task. This is congruent with reports and observations from the OR, where task demand determines the distractiveness of the soundscape. At this point, I would like to clarify that load refers to the individuals cognitive capacities to perform a task, while demand refers to the requirements a task has on cognitive capacities. For example, a surgical task can have a high demand, however, a skilled surgeon may experience a lower load compared to an inexperienced surgeon.

1.3 Electroencephalography

As note before, auditory distraction is a cognitive process that can be difficult to capture in the OR as its effects on performance are not always clear. However, a distractor must first be processed in order to be distracting. By examining auditory processing, we can investigate how the processing of the OR soundscape relates to its perception and performance during a task. To achieve this, we used mobile EEG to measure auditory processing during dynamic and surgical tasks.

The OR soundscape is characterized by a multitude of different, overlapping sounds. These encompass transient sounds like, alarms, monitor beeps or the clattering of instruments, and continuous sounds like ventilation and speech. To examine the processing of the different sounds, we applied two analytic methods, namely event-related-potentials (ERP) and temporal-response-functions (TRF). In all our studies we used a combination of both, as each method has advantages and disadvantages, which are outlined below.

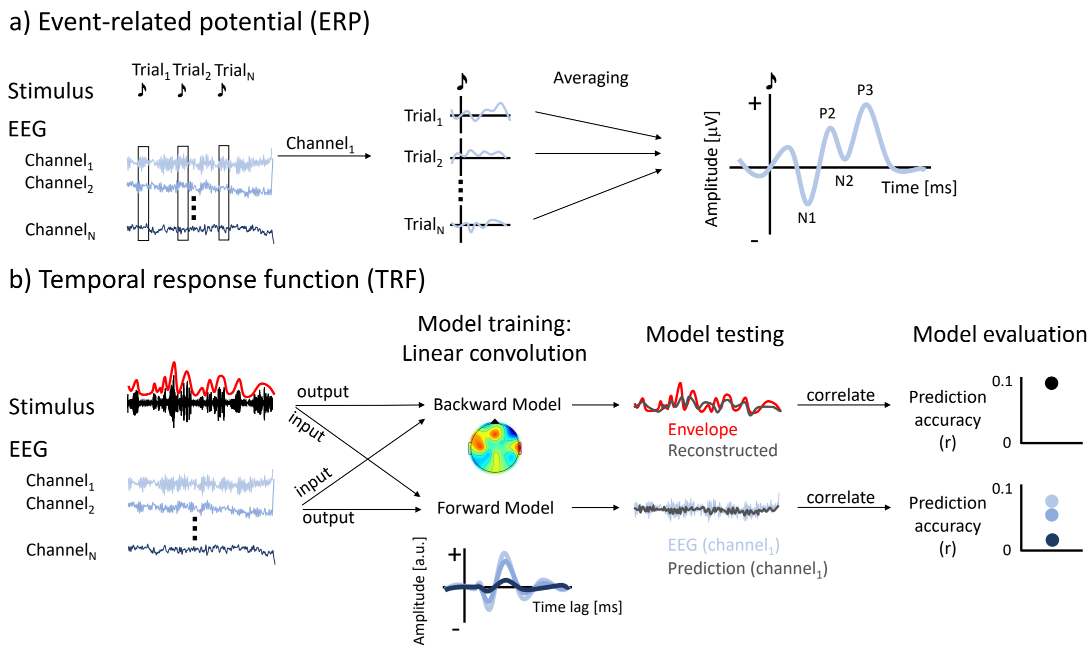


Fig. 1.1: EEG markers of auditory processing. **a)** An event-related potential (ERP) is the brain's time-locked response to a specific event, such as a sound. Due to the low signal-to-noise ratio (SNR) of single-trial ERPs, multiple trials are averaged to enhance the signal. This results in a time-course with distinct peaks that reflect different stages of auditory processing. **b)** A temporal response function (TRF) describes the relationship between a continuous stimulus, typically expressed as a feature such as the envelope, and the continuous neural response. TRFs can be computed using a backward model, which reconstructs the stimulus from the neural response, or a forward model, which maps the stimulus onto each recorded EEG channel. The backward model optimizes performance by weighting available information across channels and time points but does not yield a neurophysiologically interpretable response. In contrast, the forward model provides a time-course similar to the ERP and offers better physiological interpretability. Model accuracy can be evaluated by correlating either the reconstructed envelope with the actual envelope (backward model) or the predicted EEG signal with the recorded EEG signal for each channel (forward model).

Event-related-potentials (ERP) An ERP is the time-locked EEG response to an event that occurred externally to a person, such as an auditory or visual stimulus, or internally, such as a motor response or an eye-blink. EEG activity unrelated to the event, such as artifacts or spontaneous oscillatory activity, is regarded as noise and will be reduced by averaging across multiple repetitions of the event (Figure 1.1a; Luck, 2014). The resulting ERP time-course reflects event processing with respect to the onset of the event. The ERP can be divided into several peaks which reflect different stages of sensory and cognitive processing. The differences between these processes are expressed in varying ERP features, such as amplitude, latency and topography.

The N1 (or N100) response and the P2 (or P200) response reflect early sensory processes. The N1 is a negative deflection that typically reaches its maximum amplitude around 100 ms after stimulus onset, while the P2 is a positive deflection that peaks around 200 ms after stimulus onset. Both ERP components have been studied in a variety of context, for example how we adapt to repeated stimulus presentation (Näätänen & Picton, 1987; Ross & Tremblay, 2009). The components can also indicate a persons ability to filter out irrelevant sounds, a process called sensory gating (Lijffijt et al., 2009). Furthermore, the components are influenced by top-down processes such as attention (Näätänen & Picton, 1987), reflecting differences in attentional engagement with a stimulus.

The N2 (or N200) response and P3 (or P300) response reflect cognitive processes. The N2 is a negative deflection around 200 ms after stimulus onset and occurs after the P2. It is involved in a variety of processes including stimulus identification and attention (Patel & Azzam, 2005; Picton, Alain, Otten, Ritter, & Achim, 2000). The P3 is a positive deflection that typically starts around 300 ms after stimulus onset. The P3 has been used in the study of attention, as it can reflect goal-directed (top-down) or stimulus-driven (bottom-up) attention (Patel & Azzam, 2005; Polich, 2007).

An advantage of ERPs is that they have been extensively studied over several decades. Thus, investigating ERPs outside of traditional lab environments allows a direct comparison to a large body of literature. The first mobile EEG studies were in fact ERP studies using auditory stimuli (Debener et al., 2012; De Vos, Gandras, & Debener, 2014). Since then, auditory ERPs have been employed to examine perceptual and cognitive processes in various applied settings, including driving (Protzak & Gramann, 2018), aviation (Dehais, Somon, Mullen, & Callan, 2021), and also surgical simulations (Thomaschewski et al., 2021; Zander et al., 2017). In studies employing auditory ERPs to examine cognitive processes, artificial stimuli are commonly used as alterations in stimulus parameters, such as sound intensity, can impact the ERP peaks (Näätänen & Picton, 1987). However, presenting such stimuli in a workplace setting, such as an OR, would be unethical as it may introduce an unnecessary distraction that could negatively impact the patient's health. The ideal scenario would be to utilize the natural soundscape as an informative event, thereby revealing the cognitive and perceptual processes of the personnel. While repetitive artificial stimuli are rarely encountered in real-life, in the OR environment such repetitions occur, for example the beep of an ECG monitor. Thus, certain sounds in the OR may in fact be suitable for ERP analysis.

However, the OR soundscape also has continuous aspects, where many non-repetitive sounds are embedded in a continuous stream of auditory information.

In conclusion, studying ERPs to artificial sounds has significantly contributed to our understanding of sensory and cognitive processes within work environments. They are a valuable tool when transitioning from laboratory to real-world research. However, they are constrained by the necessary repetition of a stimulus, thus neglecting the continuous aspects of soundscapes.

Temporal-response-function (TRF) The limitation of ERPs - that a stimulus must be repeatedly presented - has been addressed by modeling continuous stimulus-response relationships, thereby enabling the study of naturalistic stimuli, like speech (Aiken & Picton, 2008; Ding & Simon, 2012b). One such approach applies a linear convolution, the TRF, either as a forward model, where the stimulus predicts the response, or as a backward model, where the response predicts the stimulus (Figure 1.1b; Crosse, Di Liberto, Bednar, & Lalor, 2016; Haufe et al., 2014). In the case of the forward model, the stimulus is mapped onto every channel of the response. The resulting time-course is similar to an ERP, containing multiple peaks with varying topographic distributions that are physiologically interpretable (Jaeger, Mirkovic, Bleichner, & Debener, 2020; Kong, Mullangi, & Ding, 2014; Petersen, Wöstmann, Obleser, & Lunner, 2017). In contrast to the forward model, the backward model reconstructs the stimulus using the response. This approach has the advantage of combining all information from the EEG (i.e. channel and time-points), with the data weighted according to their relevance. However, the resulting model is not physiologically interpretable (Haufe et al., 2014). The reliability of a forward or backward model is evaluated by correlating the actual signal with the signal predicted by the model. A high correlation value indicates a good prediction accuracy of the model.

In the field of continuous stimulus-response mapping there is a focus on speech as it is a natural, continuous stimulus that is an essential part of most people's everyday life, therefore involved in many cognitive processes (Hamilton & Huth, 2020). For example, attended speech shows larger prediction accuracies than ignored speech (e.g. Mirkovic, Debener, Jaeger, & De Vos, 2015; O'Sullivan et al., 2015). Moreover, prediction accuracy is reduced during dual-tasking (Xie, Brodbeck, & Chandrasekaran, 2023) or when attention is directed to a non-verbal task, rendering speech irrelevant (Vanthornhout, Decruy, & Francart, 2019). Therefore, attention influences the mapping of the stimulus-response relationship of speech.

Using TRFs in work environments enables the study of natural soundscapes without requiring repeated stimulus presentations. However, as this method is still emerging, further methodological investigation is needed to ensure its reliability and validity in applied settings. First, it is unclear whether reliable TRFs can be computed when a task is being performed simultaneously. While TRFs can be computed in mobile settings, for example while walking (e.g. Straetmans, Adiloglu, & Debener, 2024; Straetmans, Holtze, Debener, Jaeger, & Mirkovic, 2021), Vanthornhout et al. (2019) were unable to compute meaningful responses to irrelevant speech while participants played the game Tetris. They argued that movement during the task may have introduced artifacts, thereby reducing the signal quality to a degree that meaningful responses could not be obtained. Second, TRFs could reveal the general processing of soundscapes like the OR, where transient sounds are embedded in a constant stream of auditory information. However, there is currently little research on how real-world soundscapes apart from speech are processed (e.g. Huang & Elhilali, 2020; Lorenzi et al., 2023). Third, the computation of TRFs requires the extraction of a stimulus feature from the raw audio. In speech studies, a common feature is the envelope of the auditory signal (Crosse et al., 2016). However, for recordings within the OR, the anonymity of the patient and surgeons plays a crucial role - even more so than in other studies. Therefore, the optimal stimulus feature must provide a minimal invasion of privacy while still providing meaningful responses. In conclusion, before TRFs can be used to improve our understanding of how real-world, continuous soundscapes such as those found in the OR, are processed, we need to determine whether meaningful responses in such complex environments can be obtained.

1.4 Thesis objectives

The overarching goal of this thesis was to understand auditory distraction in the OR. To this end, we investigated whether EEG is suitable to continuously measure a person's auditory processing in such work environment. The thesis had three main aims. First, we were interested in the methods that could be applied to measure responses to natural soundscapes outside the classical laboratory setting. Second, we were interested how traditional laboratory-based findings of auditory processing and distraction can be translated to an OR environment. Third, we investigated how factors such as task demand influence the processing of naturalistic and complex soundscapes.

In the first study, we investigated the influence of attention on auditory processing during the performance of an audio-visual-motor task. We designed the experiment to parallel conditions typically encountered in a surgical environment: The task comprised a visual, dynamic, and bi-manual component. The soundscape consisted of relevant and irrelevant speech and sounds, embedded within a continuous OR playback. We evaluated the ERP N1 and P3 responses to relevant and irrelevant stimuli, providing a comparison with more controlled studies. We further computed meaningful TRFs to the continuous soundscape, thereby showing that responses to soundscapes beyond speech can be obtained while a task is being performed.

Results of the first study indicate that reliable responses to an OR soundscape can be obtained. Nevertheless, it was unclear whether these responses could provide insight into an individual's state. Therefore, in the second study, we examined the impact of task demand on the processing and perception of a soundscape. To bridge laboratory research and more naturalistic settings, we combined a simulated surgical task with a cognitive manipulation inspired by the duplex-mechanism account of distraction. Participants performed a serial recall task before the surgical task, and were presented with an OR playback in combination with a changing-state sequence, which should serve as a potentially distracting stimulus. We computed ERP N1, P2 and N2 responses, and their equivalents for TRFs to measure the processing of different aspects of the soundscape, namely transient sounds (i.e. the changing-state sequence) and the OR playback in general. Additionally, we sought to determine the amount of information about the OR playback that is required to compute meaningful responses. To this end, we computed prediction accuracies from different acoustic features of the OR playback.

While the paradigm of the second study provided a further step towards a realistic OR environment, the demand manipulation and the changing-state sequence were still artificial. In the third study, we thus sought to enhance the realism of the paradigm. The aim of this study was again to investigate the impact of demand on the processing of the soundscape. This time, we manipulated task demand by using two surgical tasks of varying difficulty to simulate surgical phases of varying demand. Furthermore, natural speech was presented within an OR playback, providing a realistic scenario of potentially distracting irrelevant speech. Again, we computed ERP N1 and P2 responses to distinct events, and calculated prediction accuracies for the different continuous stimuli of the OR environment, namely the non-speech soundscape and speech. This way, we could investigate how demand influences the processing of different aspects of the environment.

2 Study I - Investigating the attentional focus to workplace-related soundscapes in a complex audio-visual-motor task using EEG.

This article has been published as:

Rosenkranz M., Cetin T., Uslar V.N. and Bleichner M.G. (2023). Investigating the attentional focus to workplace-related soundscapes in a complex audio-visual-motor task using EEG. *Front. Neuroergon.*. 3:1062227. doi: 10.3389/fnrgo.2022.1062227

Author contributions: MR and MB conceptualized the experiment. MR performed the data acquisition, analyzed the data, and wrote the manuscript to which TC, VU, and MB contributed with critical revisions. All authors approved the final version and agreed to be accountable for this work.

2.1 Abstract

Introduction: In demanding work situations (e.g., during a surgery), the processing of complex soundscapes varies over time and can be a burden for medical personnel. Here we study, using mobile electroencephalography (EEG), how humans process workplace-related soundscapes while performing a complex audio-visual-motor task (3D Tetris). Specifically, we wanted to know how the attentional focus changes the processing of the soundscape as a whole.

Method: Participants played a game of 3D Tetris in which they had to use both hands to control falling blocks. At the same time, participants listened to a complex soundscape, similar to what is found in an operating room (i.e., the sound of machinery, people talking in the background, alarm sounds, and instructions). In this within-subject design, participants had to react to instructions (e.g., "place the next block in the upper left corner") and to sounds depending on the experimental condition, either to a specific alarm sound originating from a fixed location or to a beep sound that originated from varying locations. Attention to the alarm reflected a narrow attentional focus, as it was easy to detect and

most of the soundscape could be ignored. Attention to the beep reflected a wide attentional focus, as it required the participants to monitor multiple different sound streams.

Results and discussion: Results show the robustness of the N1 and P3 event related potential response during this dynamic task with a complex auditory soundscape. Furthermore, we used temporal response functions to study auditory processing to the whole soundscape. This work is a step toward studying workplace-related sound processing in the operating room using mobile EEG.

2.2 Introduction

Auditory attention, i.e., focusing on relevant sounds and ignoring irrelevant sounds, is a fundamental skill at workplaces with both, a high level of responsibility and a soundscape containing a variety of sounds. Surgery staff, for example, performs a highly complex task while being exposed to conversations, machine and tool sounds, and background music. This soundscape accumulates to sound pressure levels regularly exceeding 50dB(A) (Baltin et al., 2020; Engelmann et al., 2014; Hasfeldt, Laerkner, & Birkelund, 2010; Tsiou et al., 2008). The soundscape can become a burden for the medical staff (Healey et al., 2007; Jung, Jüni, Lebovic, & Grantcharov, 2020; Maier-Hein et al., 2022; Pleban, Radosz, Kryst, & Surgiewicz, 2021; Tsiou et al., 2008; van Harten et al., 2021), and increases surgical complication rates (Engelmann et al., 2014; Kurmann et al., 2011). Interestingly, the focus of attention to sounds and their interpretation changes throughout a surgery. For example, conversations of others are sometimes perceived as disturbing when concentration is high, while other times they are attended to and even encouraged (van Harten et al., 2021). In the former case auditory attention is focused only on task-relevant sounds (e.g., instructions or alarm sounds) and suppresses irrelevant sounds (e.g., chatting). In other words, the attentional focus is narrowed to the task. In the latter case attention switches between multiple sound sources, such as task-relevant instructions and task-irrelevant chatting. In other words, the attentional focus is wide and a large extend of the soundscape is processed. Our goal was to study a narrow compared to a wide focus to better understand auditory attention in a complex and multi-sensory environment.

Electroencephalography (EEG) can be used to measure auditory attention continuously, objectively, and without the interruption of a person. Mobile EEG systems allow to study the brain in a working environment rather than in the lab (Hölle, Meekes, & Bleichner, 2021; Wascher et al., 2014) and has already been used to assess performance during laparoscopic training and simulation (Maier-Hein et al., 2022; Pugh et al., 2020; Shafiei et al., 2021; Suárez et al., 2022; Thomaschewski et al., 2021). Thus, with EEG we want to study auditory attention in the operating room and understand when sounds become a burden.

When transitioning from the lab to the operating room, we must consider that our expectation about auditory attention is mainly derived from highly controlled studies (Gramann et al., 2021). In order to generalize lab findings to more complex environments we have to increase the environmental complexity. One approach to increase complexity is naturalistic laboratory research, which provides a balance between stimulus control and ecological validity (Matusz et al., 2019). We decided to develop a complex and dynamic, audio-visual-motor task while maintaining experimental control over stimuli. Thereby, EEG responses related to auditory attention can be studied in a complex environment.

We first operationalized the soundscape of an operating room into five stimulus categories: a continuous background stream, as well as, task relevant and irrelevant sounds, and task relevant and irrelevant speech stimuli (Engelmann et al., 2014; Hasfeldt et al., 2010). The background stream represents sounds originating from running machines, ventilation, and people moving around. Task relevant speech represents exchanges about the surgery, as well as, instructions. Task irrelevant speech represents private conversations. Task relevant sounds represent, e.g., alarm sounds and feedback from instruments. Task irrelevant sounds represent, e.g., phone ringing or sounds from monitors.

We then combined our operationalization of the soundscape with a visual-motor task, namely the computer game Tetris. The game requires the use of both hands to navigate blocks. For the continuous background stream, we chose a hospital soundscape. For task relevant speech, participants received instructions within the game. For task irrelevant speech a conversation unrelated to the game was presented. The task relevant sound changed between two conditions. For task irrelevant sounds, monitor sounds from a surgery machine were presented.

Lastly, we manipulated the attentional focus of the participants by changing the task relevant sound while keeping the complexity of the soundscape constant. In a narrow attentional focus condition (from here on narrow condition) participants had to attend to an alarm sound (from here on the alarm). This sound originated from a specific location, i.e., was easy to detect. The rest of the soundscape (except the task relevant speech) could be ignored. In a wide attentional focus condition (from here on wide condition) we implicitly direct the participants attention toward all sound streams. This was approached by instructing participants to attend to a sound that was embedded in any of the five streams. We refer to this sound as the beep, as it served the purpose of manipulating participants attention but was generally unrelated to the operating room soundscape.

Our study addressed two research questions: First, can we investigate well-known EEG responses, namely event-related potentials (ERPs) and temporal response functions (TRFs) in a dynamic task with a complex soundscape using a mobile EEG setup? Second, what are the differences in neural processing when the auditory attentional focus was narrow (i.e., most of the soundscape can be ignored) compared to wide (i.e., most of the soundscape must be attended to)? We used ERPs to study responses to distinct stimuli, i.e., relevant and irrelevant sounds, and focused on two components, the N1 and P3.: The N1 is an early negative deflection related to auditory processes and modulated by attention (Hansen & Hillyard, 1980; Hillyard, Hink, Schwent, & Picton, 1973; Luck, 2014; Picton & Hillyard, 1974). For our first hypothesis, we expected a larger N1 for irrelevant sounds (i.e., non-target sounds in both conditions) in the wide condition than in the narrow condition. Attention to the beep, which was integrated into other sounds, should lead to a stronger processing of the whole soundscape. Therefore, we expected a stronger processing of the irrelevant sounds.

The P3 is a late positive deflection in response to target sounds (from here on targets). As this response is absent in non-targets, it thereby marks attentional processes (Luck, 2014; Polich, 2007). For our second and third hypotheses we expected a P3 to the target of the respective conditions. The alarm was the target in the narrow condition, thus, we expected a larger response in this condition compared to the wide condition. The beep was the target in the wide condition, thus, we expected a larger response in this condition compared to the narrow condition.

We used TRFs to investigate processing of the soundscape as a whole. TRFs are the result of correlating a continuous EEG signal with a continuous audio signal (Crosse et al., 2016). The correlation (i.e., response) is larger for attended compared to unattended signals (Mirkovic et al., 2015; O'Sullivan et al., 2015). For our fourth hypothesis, we expected a larger TRF in the wide compared to the narrow condition, as the beep should direct attention toward the whole sound environment.

2.3 Methods

This study was registered prior to any human observation of the data (https://osf.io/sgvk6). Deviations from our preregistration are described in the supplementary material. We provided the experiment, as well as the code and data to reproduce the statistical analyses and figures here: Rosenkranz and Bleichner (2022).

2.3.1 Participants

Twenty-two participants (age range: 20-30 years; female: 16) were recruited through an online announcement on the University board. We based the sample size on previous studies showing P3 effects in naturalistic settings (e.g., Hölle et al., 2021; Protzak & Gramann, 2018; Scanlon, Townsend, Cormier, Kuziek, & Mathewson, 2019) due to the exploratory approach of this study. All participants signed prior to the experiment informed consent approved by the medical ethics committee of the University of Oldenburg and received monetary reimbursement. Eligibility criteria included: Normal hearing (self reported), normal or corrected vision, no psychological or neurological condition, right-handedness, and compliance with current COVID-related hygiene regulations (e.g., this could include proof of vaccination).
Two participants were excluded from the final analysis. One participant showed high impedance ($>$100 kΩ) for 10 channels at the end of the experiment and overall poor data

quality. One participant had a very low hit rate which indicates that this participant did not follow task instructions. The final sample consisted of 20 participants (female: 14).

2.3.2 Paradigm

Participants performed a complex audio-visual-motor task - an adapted 3D Tetris game. The basis for the game was developed by Kalarus (2021) and we changed it to our needs. Below is a short description of the paradigm. For a detailed description of the game and generation of the auditory stimulus material see Supplementary material.

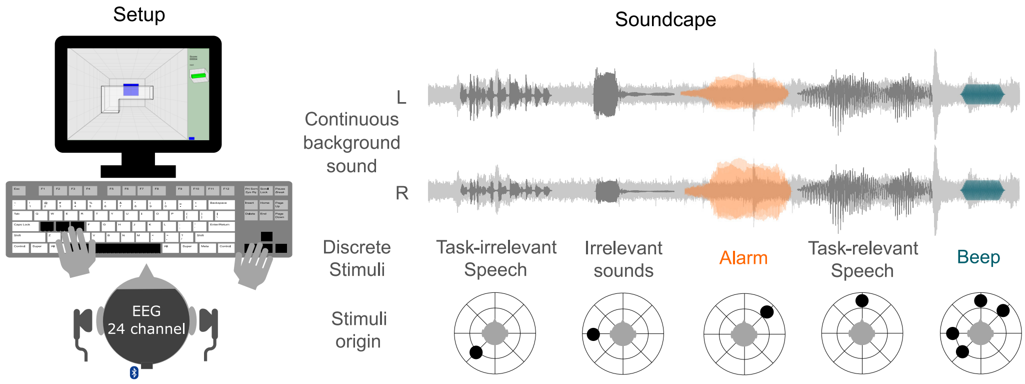


Fig. 2.1: A) Experimental Setup: Participants played 3D Tetris (with the left hand participant controlled the rotation on a block, with the right hand the position of a block). The soundscape was presented via headphones. EEG was recorded using a 24 channel mobile EEG setup. **B)** Soundscape: A continuous background sound was presented throughout the task. Discrete stimuli were subsequently presented. The alarm was the target in the narrow condition, while the beep was the target in the wide condition. The alarm was presented from one direction, the beep was presented from any direction as the other sounds. If participants detected a target, they should press the space bar.

Participants had to play a 3D Tetris game while reacting to different sounds and instructions (see Figure 2.1A). In 3D Tetris one is presented with a three-dimensional space in which differently shaped, three-dimensional blocks are placed. The falling blocks must be placed in such a way that they form a layer. When a layer is complete, the layer disappears. Participants controlled the rotation of the blocks with the left hand and position of the blocks with the right hand. The goal was to place blocks to remove as many layers as possible to receive points. Unlike the classic Tetris game, participants could not loose when the blocks were stacked too high. Instead, the game restarted at the bottom layer to allow for a continuous game-play. When that happened, participants lost points.

Furthermore, participants were listening to a soundscape. The soundscape included one continuous background sound, and five discrete stimuli (see Figure 2.1B). The background sound consisted of hospital sounds, e.g., air conditioning and people moving around, and was presented from both sides. A task irrelevant speech of two people talking in the background originated to the left behind the participant (-135∘). Two irrelevant sounds were presented from the left side (-90∘). Participants also received instructions from time to time from the front (0∘) on where to place the next block. Furthermore, an alarm was presented from the right (45∘) and a beep could occur from the same direction as the other stimuli. All sounds were spatially separated using the Head Related Impulse function (Kayser et al., 2009).

For the auditory task, participants should have attend to the task relevant speech, which instructed participants to place the next block in one of the four corners of the Tetris layer. Furthermore, participants played the game twice (a game lasted approximately 18 min) and received a different instruction for each condition. Note, that the soundscape was conceptually the same for both conditions. In the narrow condition, participants were instructed to additionally attend to the alarm, i.e., participants had to attend to the task relevant speech and the alarm. The alarm was long, had a high intensity, and was presented always from the same direction, thus, it was not necessary to attend to the rest of the soundscape to detect it. In the wide condition, participants were instructed to attend to the beep, i.e., participants had to attend to the task relevant speech and the beep. The beep was short, had a low intensity, and was integrated into other stimuli, thus, the whole soundscape had to be monitored to detect it. To summarize, the difference between the two conditions were the instruction on which target should be attended to. The target of the narrow and wide condition were the alarm and the beep, respectively. If participants detected a target, they should press the space bar. Hitting a target, as well as, following the speech instructions, granted points, while misses and not following instructions subtracted points.

All discrete auditory stimuli were initially presented 48 times in a random order. However, the response to the beep overlapped with the response to the alarm and irrelevant sounds when it was integrated into them. Therefore, we added all overlapping sounds again to derive at 48 non-overlapping sounds. Note, that only responses to non-overlapping sounds were used in the ERP-analyses.

To get acquainted with the game, participants received written instructions for the game. Then, a general training without auditory stimuli and a training for the relevant speech instructions was performed. Before each condition, participants also performed a condition specific training in which they received feedback on whether they correctly detected the target (see Supplementary material for a detailed description of the training games). EEG was not recorded during the training games. Before the game of each condition started, resting EEG was recorded, by instructing participants to first focus on a fixation cross and then close their eyes for 1 min each. Furthermore, two questionnaires were administered: At the beginning of the experiment, participants filled out a noise sensitivity questionnaire (NoiSeQ—results are not part of the current study; Schutte, Marks, Wenning, & Griefahn, 2007) and after each condition a workload questionnaire (NASA-TLX; Hart & Staveland, 1988).

2.3.3 Data acquisition

Participants were asked to wash their hair on the day of recording. EEG data was recorded using a wireless 24channel amplifier (SMARTING, mBrainTrain, Belgrade, Serbia) attached to the back of the EEG cap (EasyCap GmbH, Hersching, Germany) with Ag/AgCl passive electrodes (see Supplementary Figure S2.2 for the channel layout) and the reference and ground electrode at position Fz and AFz, respectively. The data was recorded using a sampling rate of 500 Hz, and transmitted via Bluetooth from the amplifier to a Bluetooth dongle (BlueSoleil) that was plugged into a computer (Dell Optiplex 5070).

After fitting the cap, the skin was cleaned using 70% alcohol. To increase skin conductance between the scalp and electrodes abrasive gel (Abralyt HiCl, Easycap GmbH, Germany) was used. Impedance were kept below 20 kΩ at the beginning and again checked at the end of the recording using the SMARTING Streamer software (v3.4.3; mBrainTrain, Belgrade, Serbia). Recording took place in a quite and electrically shielded room. Participants were seated in front of a screen (Samsung, SyncMaster P2470) and keyboard (Dell, KB 1421). Auditory stimuli were presented using Psychtoolbox 3 (v3.0.17, Kleiner et al., 2007). For each stimulus type, a sound marker was generated using the lab streaming layer library[1]. A key capture software [2] was used to record which key was pressed on the keyboard and an audio capture software (this is used as input for the TRF analysis, see below) [3] was used to record the presented audio with a sampling rate of 44100 Hz. To synchronize all data streams, the transmitted EEG data, sound marker, keyboard marker, and computer audio was collected in the Lab Recorder software [4] based on the Lab Streaming Layer and saved as .xdf files. The same computer was used for data recording and experiment presentation.

2.3.4 Preprocessing

The EEG was analyzed using EEGLAB (v2021.0, Delorme & Makeig, 2004) in MATLAB R2020b (The MathWorks, Natick, MA, United States).

For each participant and condition, the continuous data was filtered with Hamming windowed FIR filter using the EEGLAB default settings: (1) high-pass: passband edge = 0.1 Hz[5]; (2) low-pass: passband edge = 30 Hz[6]. These filter settings are recommended for ERP analyses (Luck, 2014), as well as TRF analyses (Crosse et al., 2021). The filtered data was re-sampled to 250 Hz. Channels were visually checked for flat lines and bad data quality (e.g., if impedance were above 20 kΩ). Bad channels were removed from both conditions. Afterwards, the data was cleaned from artifacts using infomax independent component analysis (ICA) and rejected channel were interpolated.

[1] https://github.com/labstreaminglayer/liblsl-Matlab, v1.14
[2] https://github.com/labstreaminglayer/App-Input, v1.15
[3] https://github.com/labstreaminglayer/App-AudioCapture, v1.14
[4] https://github.com/labstreaminglayer/App-LabRecorder, v1.14
[5] filter order = 16500, transition bandwidth = 0.1 Hz, cutoff frequency (-6dB) = 0.05 Hz
[6] filter order = 220, transition bandwidth = 7.5 Hz, cutoff frequency (-6dB) = 33.75 Hz

For the ICA, a copy of the preprocessed data was high-pass filtered (passband edge = 1 Hz[7]) and cut into consecutive epochs of one second. Epochs with a global or local threshold of 2 standard deviations were automatically rejected. ICA decomposition was applied on the remaining epochs of both conditions. The resulting components were back-projected on the original preprocessed, but uncleaned data of each condition. Components related to eye-blinks, eye-movement, heart rate, and muscle movement were identified and removed using the EEGLAB toolbox ICLabel (Pion-Tonachini, Kreutz-Delgado, & Makeig, 2019) with a threshold of .9. On average, 2.8 ($\pm$ 1.32) components were rejected. Afterwards, previously rejected channels were interpolated using spherical interpolation. Then, channel were re-referenced to the linked mastoids (TP9/TP10). For all auditory stimuli, a constant delay of 19 ms between the sound marker and sound presentation was taken into account.

ERP analysis

ERP analyses were performed for the alarm, beep, and the two irrelevant sounds. For each of the sounds, epochs from -200 to 800 ms with respect to the stimulus onset were generated and a baseline correction from -200 to 0 ms prior to stimulus onset was performed. Epochs with a global or local threshold of 3 standard deviations were automatically rejected. For targets (i.e., the alarm or beep), only hit trials were included in the analysis. A hit was defined as any space bar press within 3 seconds after a target.

We calculated ERP amplitudes averaged over time based on individual time-windows. Our ERP analyses focused on the N1 and P3 component. The analyses of the two components were identical, except for the selection of channel and time-window. For each participant, an average response was calculated from the two conditions and selected channels. The ERP N1 is typically associated with a negative frontal polarity around 100 ms after stimulus presentation (Näätänen & Picton, 1987) and the ERP P3 with a positive parietal polarity around 300 ms after stimulus presentation (Polich, 2007). For the N1, we selected channel Fz, FC1, FC2, Cz, C3, and C4; and for the P3, we selected channel Pz, P3, P4, CPz, CP1, and CP2 (see Supplementary Figure S2.2). The average response was used to find the component peaks of each participant. For the N1, we searched for a negative deflection between 50 and 150 ms following stimulus onset. For the P3, we searched for a positive deflection between 300 and 400 ms following stimulus onset. Following peak detection, the component time-window was determined. For this, a time-window of ±25 ms and ±50 ms around the N1 and P3 peak was taken, respectively. Lastly, to derive at trial-level data, for each participant, condition, selected channel, and trial, the mean amplitude over the individual time-window was calculated.

[7]filter order = 1650, transition bandwidth = 1 Hz, cutoff frequency (-6dB) = 0.5 Hz

TRF analysis

For the analysis of the soundscape as a whole, the mTRF toolbox (Crosse et al., 2016) was used. Therefore, the recorded audio was preprocessed as follows: First, the absolute of the Hilbert transform of the audio was low-pass filtered at 30 Hz[8] and resampled to 250 Hz. Second, we were interested in the response to the whole soundscape irrespective of the targets, therefore, the alarm and beep were excluded from the EEG and audio data of both conditions. Data from the onset of the alarm and beep was were excluded up to 1 s after the onset, creating epochs of unequal length. The total length of data was unequal between conditions, therefore we excluded epochs until the total length difference between conditions for each participant was $<$1 min. Third, EEG data was multiplied by factor 0.0313 for normalization (as suggested in the provided code by (Crosse et al., 2016)). Finally, a forward model was trained on the epoched EEG data and audio data using the function *mTRFtrain*. Time lags were calculated from -200 to 800 ms and a lambda of 0.1 was used.

The TRF usually reveals classic ERP peaks known from the auditory processing literature (Crosse et al., 2016; Jaeger et al., 2020; Mirkovic et al., 2015). Based on pilot data from three participants (not included in the final analyses), we expected these peaks at approximately 100, 200, and 300 ms time lag.

We verified these condition-independent peaks using a permutation-based approach, which was implemented with the Mass Univariate ERP Toolbox (Groppe, Urbach, & Kutas, 2011). First, the TRF of each participant, condition, and channel was baseline corrected within the function *sets2GND* using time lags from -100 to 0 ms. Second, two-sided t-values were calculated and corrected for multiple comparisons within the function *tmaxGND* using a time-window from 0 to 450 ms time lag. Finally, a time-window was identified as significant when t-values exceeded a significant threshold of $p < 0.05$.

Within the significant time-windows, we determined individual TRF peaks. For this, we first calculated the standard deviation over channels to derive the global field power (GFP) of the TRF. The GFP indicates the magnitude of a signal across channels. Thereby, it accounts for individual differences in spatial distribution and avoids channel selection (Murray, Brunet, & Michel, 2008). The resulting GFP of each condition were averaged. Next, we searched for the condition-averaged, maximum GFP value in each significant time-window and for each participant. Then, we calculated for each participant the full width at half maximum with respect to the peak to determine individual time-windows. Finally, we averaged over the individual time-windows of the GFP of each condition. This resulted in an average GFP value for each participant, condition, and significant time-window.

[8] filter order = 220, transition bandwidth = 7.5 Hz, cutoff frequency (-6dB) = 33.75 Hz

2.3.5 Statistics

Preregistered Analyses

Condition differences of the auditory task were analyzed using a linear mixed model (LMM). The analysis was performed in RStudio (version 2021.09.0) using the R package lmer4 (version 1.1-23). For all analyses a categorical fixed factor 'condition' with two categories was used, i.e., narrow and wide, which were coded 0 and 1, respectively. For the ERP analysis, the response amplitude was predicted for each trial. Participant and channel were included as random factors (Volpert-Esmond, Page-Gould, & Bartholow, 2021):

$$AMP \sim condition + (condition|participant) + (1|channel) \tag{2.1}$$

For the ERP model of the beep, we encountered singularity issues. The random factor for channel showed a variance of 0 indicating over-specification of this random factor (Volpert-Esmond et al., 2021). We therefore excluded this factor when computing the model for the beep.

For the TRF analysis, GFP differences were predicted for individual time averaged peaks. Participants were included as a random factor:

$$GFP \sim condition + (1|participant) \tag{2.2}$$

LMMs allow the investigation of the random factors participant and channel. For this, the intraclass correlation coefficient (ICC) was used, which represents the amount of variance in the predicted variable that is explained by the random factors (Lorah, 2018; Volpert-Esmond et al., 2021). Variances for each factor were calculated using an intercept only model for the analysis of ERPs ($AMP \sim 1 + (1|participant) + (1|channel)$) and TRFs ($GFP \sim 1 + (1|participant)$). ICCs were calculated by dividing the variance of participant or channel by the total variance.

Fixed effects were evaluated using Satterthwaite approximations within the R package lmerTest, which estimates the degrees of freedom to calculate two-tailed p-values. Evidence for an effect were assumed for p-values below 0.05. We also report standard errors (SE) and 95

Exploratory Analyses

For a better understanding of the performance on the Tetris and auditory task, we explored results of the NASA-TLX and several behavioral measures. The NASA-TLX was used to investigate differences in perceived workload between conditions. The questionnaire has six subscales, with scores ranging from 0 to 20. A high score is associated with high workload. We summed the scores of all subscales to receive one workload score per condition and participant (i.e., scores can range from 0 to 120; Hart & Staveland, 1988). We checked task performance on the Tetris task by comparing the number of completed layers. We also checked performance on the speech instruction task by comparing the number of instructions that were correctly followed. Workload and behavioral scores were compared between conditions by computing LMMs for each score (i.e., workload; completed layers; followed instructions) with condition as a fixed factor, and participant as a random factor ($Score \sim condition + (1|participant)$).

We further explored reaction time and hit rate in response to the targets, i.e., the target in the narrow condition was the alarm and in the wide condition the beep. For this we used all trial-level responses, i.e., also those that were not considered in the ERP-analyses.

Individual reaction times between conditions were compared using a generalized linear mixed model (GLMM) with an inverse Gaussian distribution to account for a positive skew in reaction time data ($Reactiontime \sim condition + (condition|participant)$).

Hit rate followed a binomial distribution, as hits and misses were coded one and zero, respectively. Differences between conditions was therefore compared using a GLMM with a binomial distribution and logit link function ($Hitrate \sim condition + (condition|partcipant)$).

The statistical significance of differences in reaction times or hit rate between the alarm in the narrow and the beep in the wide condition was evaluated using the Wald Chi-square test.

2.4 Results

2.4.1 Behavioral results

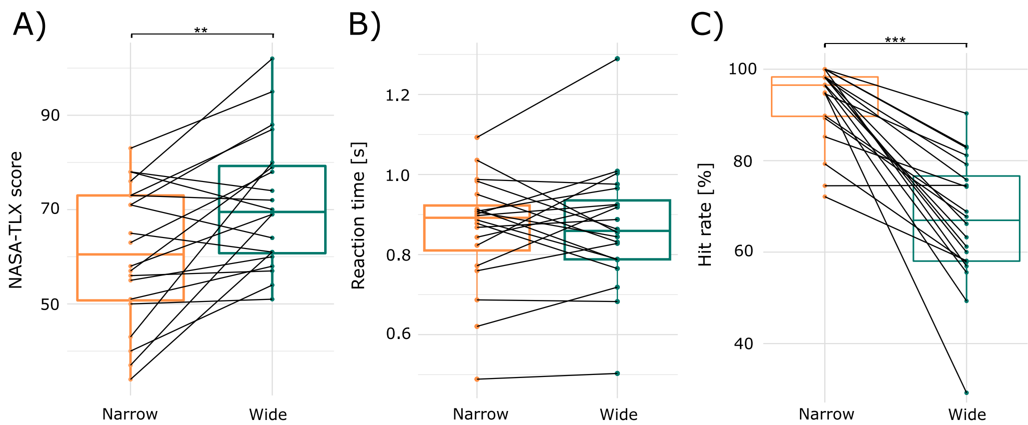


Fig. 2.2: **A)** Subjective workload scores measured by the NASA-TLX. **B)** Trial-averaged reaction time and **C)** trial-averaged hit rate in response to the targets of the respective condition, i.e., the alarm and the beep in the narrow and wide condition, respectively. Each black line represents one subject. **p<.01 ***p<.001

Figure 2.2A shows the subjective workload of each participant and condition. The average rating in the narrow condition was 60.6 which significantly increased in the wide condition to 71.45 (b=10.85; SE=3.22; p=0.0016; CI=[4.92 16.78]). There were no differences between conditions in the number of completed layers (Supplementary Figure S2.3A; b=-1.37; SE 1.09; p=0.222; CI=[-3.55 0.797]) and followed speech instructions (Supplementary Figure S2.3B; b=-0.77; SE 0.92; p=0.458; CI=[-2.55 1.15]). Figure 2.2B and C shows the performance of the auditory task, i.e., reaction time and hit rate in response to targets. Here, the response to the alarm in the narrow condition is compared to the response to the beep in the wide condition. Estimated mean reaction times in response to the alarm in the narrow condition were 0.814 seconds (b=1.22; SE=0.064) and to the beep in the wide condition 0.81 seconds. Reaction times did not differ between the two targets (Figure 2.2A; b=-0.008; SE=0.045; p=0.869). However, the chance of hitting a target in the narrow condition was 96.2% and in the wide condition 68%. The beep was significantly less often detected than the alarm (Figure 2.2B; b=-2.478; SE=0.344; p<0.001).

2.4.2 Event-related potentials

We investigated ERPs in response to task-relevant and irrelevant sounds. The alarm was relevant in the narrow and the beep relevant in the wide condition. For these sounds we investigated the P3, and expected that targets show larger P3 amplitudes than non-targets. The irrelevant sounds were ignored in both conditions. Here we investigated the N1 and expected a lower amplitude in the wide compared to the narrow condition.

The alarm Figure 2.3A shows the grand average ERP (i.e., averaged over participants and selected channel) in response to the alarm in the two conditions. We see a clear N1 peak around 100 ms, a P2 peak around 200 ms, and a P3 that starts around 300 ms. The topographies of the narrow condition shows a typical parietal P3 distribution (Polich, 2007). The mean amplitude of the P3 for the alarm in the narrow condition was 4.2 μV with a significant mean amplitude decrease in the wide condition of -2.3 μV (SE=0.53; CI=[-3.39, -1.27]; p<.001).
Computing the ICC showed that variance between people and channel accounted for 12.2% and 0.1% of the total variance, respectively (see Supplementary Table S2.2 for the results of the random effect models).

The beep Figure 2.3B shows the grand average ERP in response to the beep. For the beep we neither see a clear N1, nor a P2 peak, but a P3 that starts around 300 ms. The topography also reflects a P3 to the beep in the wide condition. The mean amplitude of the P3 for the beep in the narrow condition was -.609 μV with a significant mean increase in the wide condition of 4.1 μV (SE=0.92; CI=[2.58, 6.16]; p<.001). The ICC showed that the variance between people and channel accounted for 6% and 0% of the total variance, respectively.

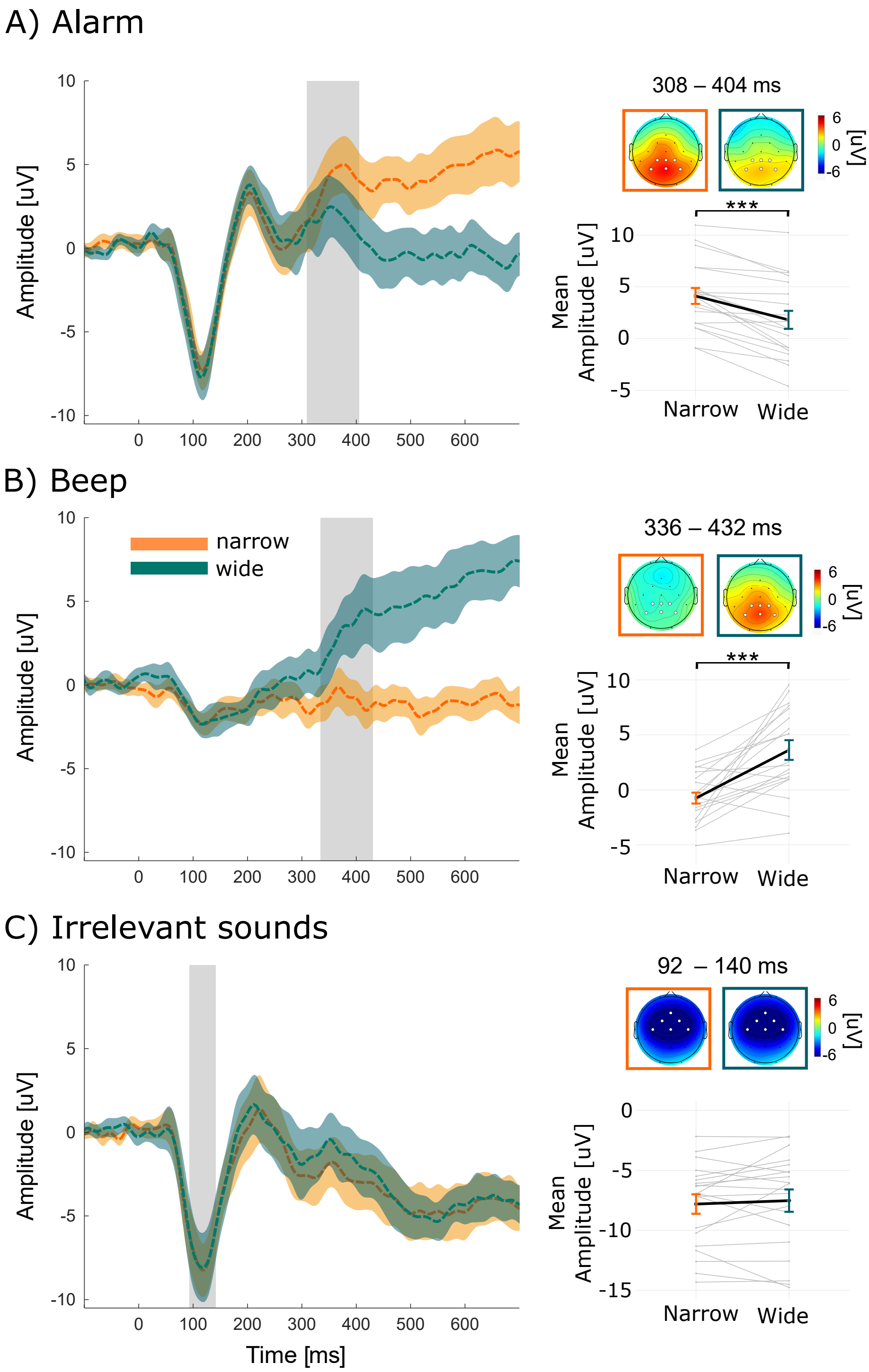


Fig. 2.3: ERPs of the **(A)** Alarm, **(B)** Beep, and **(C)** Irrelevant sounds for each condition and averaged over participants and selected channel. The selected channels are marked in white in the topographies. The narrow and wide condition are marked with orange and green, respectively. Color shades indicate the 95% confidence interval. The gray area indicates the average time-window and the topographies show the average amplitudes averaged over this time-window. Note, that individual time-windows were used for the statistical comparison. Below the topographies, the fixed effects (thick black lines) and the variability of effect between individuals (i.e., each gray line corresponds to one participant) are displayed. *** $p<.001$

Irrelevant sounds Figure 2.3C shows the grand average ERP in response to the irrelevant sounds. We see a clear N1 peak around 100 ms and a P2 peak around 200 ms. The mean amplitude of the N1 for the irrelevant sound in the narrow condition was -7.81 μV which did not differ from the wide condition (b=-0.28; SE=0.48; CI=[-1.02, 0.87]; p=0.558). The topography reflect a frontal N1 in both conditions. The ICC showed that variance between people and channel accounted for 14.8% and 0.2% of the total variance, respectively.

Summary of ERP results We found evidence for two of our three hypotheses. The alarm and beep both showed a P3 when they were the target. This shows that participants were able to detect the sounds. The P3 subsides after approximately 1.5 seconds (see Supplementary Figure S2.4). The response to the irrelevant sound did not change. In the following we look at the processing of the entire soundscape.

2.4.3 Processing of the soundscape as a whole

Figure 2.4A shows the TRF for both conditions (colored) and condition-independent (black). Condition specific and independent TRFs show a typical shape (Crosse et al., 2016). The gray areas indicate the three time-windows (0-68 ms; 96-192 ms; 216-448 ms time lag) that significantly differed from zero. The topographies show TRF values averaged over the significant time-window. All time-windows show the largest values across the fronto-central channels. For the first and last time-window the values were negative, while the second time-window showed positive values. Figure 2.4B illustrates the grand average GFP over all participants. In each significant time-window, we determined individual time-windows of the GFP and calculated the average amplitude over the individual time-window. The results are shown in Figure 2.4C. The individual GFP in the third time-window was on average 5.57 in the narrow condition, which significantly increased in the wide condition to 6.43 (b = 0.77; SE = 0.3; CI = [0.15 1.38]; p = 0.0211). We did not find significant differences for the first (b = -0.26; SE = 0.68; CI = [-1.63 1.1]; p = 0.705) and second (b = 0.11; SE = 0.59; CI = [-1.07 1.3]; p = 0.853) time-window.

The ICC of the third time-window showed that variance between people accounted for 73.3% of the total variance, indicating a large between-person variance. Note, that the high between-person variance of the TRF compared to the ERPs is the result of using averaged compared to trial-level data, respectively.

One participant showed extremely high standard deviations (see Supplementary Figure S2.8) across all channels, thus, we excluded this participant and ran the analyses again, however, this did not change the results (First: b = 0.28; SE = 0.43; CI = [-0.58 1.15]; p = 0.5174; Second: b = 0.44; SE = 0.52; CI = [-0.61 1.48]; p = 0.414; Third: b = 0.84; SE = 0.31; CI = [0.22 1.47]; p = 0.014).

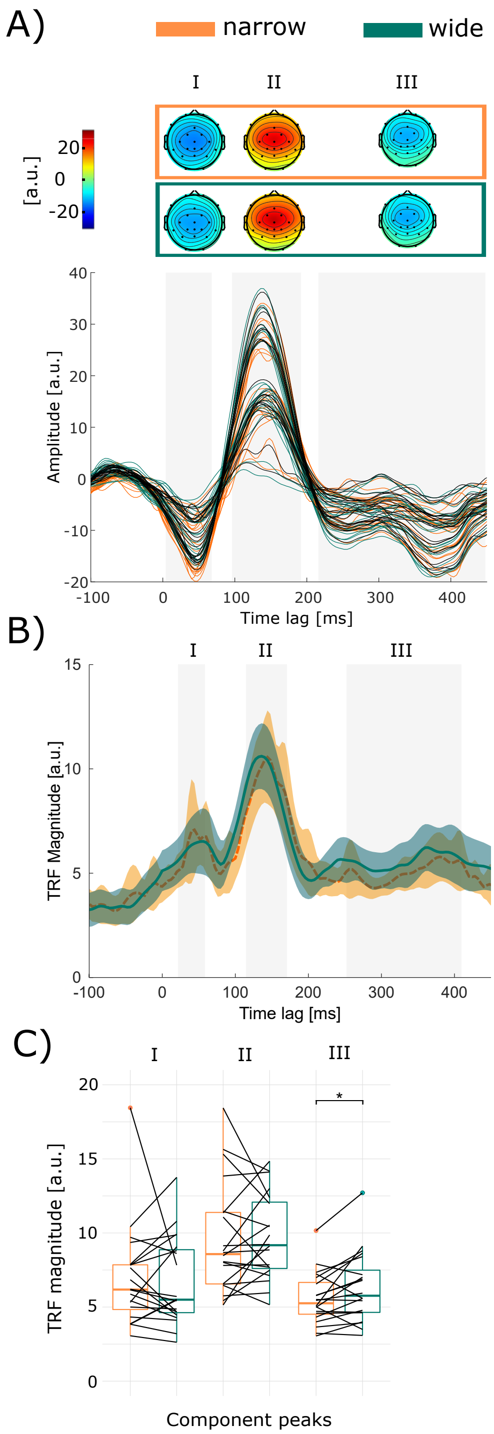


Fig. 2.4: A) The topographies show the significant time-windows averaged over time. Below are the TRFs for each channel, averaged over participants. Orange and green mark the narrow and wide condition, respectively. Black marks the TRF calculated over both conditions. The significant time-windows are marked in gray. **B)** The GFP of the TRF is shown for each condition, averaged over participants. Color shades mark the 95% confidence interval. For each participant, an individual time-window within the significant time-window was calculated. The gray area marks the average individual time-window **C)** Boxplots show the differences of the individual time-window for each significant time-window. Each line represents one participant. * $p<.05$

From Figure 2.4B and the supplementary figure S2.8, it appears that the time-windows vary largely between participants, especially with regards to the last time-window. Therefore, we re-analyzed the data by averaging over the significant time-windows that are seen in Figure 2.4A. Using the same time-window for each participant, we receive the same results (First: b = -0.03; SE = 0.46; CI = [-0.96 0.9]; p = 0.95; Second: b = 0.13; SE = 0.48; CI = [-0.84 1.1]; p = 0.791; Third: b = 0.628; SE = 0.26; CI = [0.106 1.15]; p = 0.026).

2.5 Discussion

We investigated auditory attention using a mobile EEG setup while participants completed a complex audio-visual-motor task with a rich soundscape. We manipulated auditory attention while keeping the complexity of the soundscape constant. In both conditions, participants attended to a target. In one condition, this target was a clearly audible alarm originating from one direction which required a narrow attentional focus. In the other condition, this target was a beep originating from different directions which required attention to the whole soundscape, i.e., a wide attentional focus.

Behaviorally, we found, that the sound that was assumed to be more difficult to detect (i.e., the beep), was indeed less often detected than the sound that was assumed to be easy to detect (i.e., the alarm). This is also reflected in perceived workload, which was higher in the wide condition than the narrow condition, but in contrast to Tetris and speech instruction performance which was similar across conditions. It appears that a trade-off between the Tetris task and the auditory task occurred. In other words, if the auditory task was too difficult, participants rather concentrated on the Tetris task and speech instructions.

Regrading the ERPs, we found a larger P3 if a sound was a target compared to the same sound if it was not the target, i.e., the response was larger for the alarm in the narrow compared to wide condition and for the beep in the wide compared to narrow condition. We observed the difference around 300 ms after stimulus onset. Contrary to our expectation, we did not find a clear difference in the N1 to stimuli that were irrelevant in both conditions. We also found that the TRF was larger in the last time-window for the wide compared to narrow condition.

Processing of relevant stimuli We observed a larger P3 for target compared to non-target stimuli, which indicates that participants were generally performing the auditory task. The P3 is related to attentional processed and has two subcomponents, the P3a and P3b (Polich, 2007). The P3a is typically associated with an attention switch to novel or salient stimuli and shows a central topography, while the P3b response is typically elicited by task-relevant stimuli and shows a parietal topography (Luck, 2014; Polich, 2007). In this study we focused on the P3b, as attention to different task-relevant sounds (i.e., the alarm or beep) should lead to a narrow or wide attentional focus. Our findings show the robustness of the P3b even in a complex task with visual input, auditory instructions, and motor responses.

Therefore, it lines up with beyond-the-lab studies that showed the P3b during walking (Debener et al., 2012), biking (Scanlon et al., 2019) , driving (Protzak & Gramann, 2018), and office work (Hölle et al., 2021).

Importantly, the P3b morphology was comparable between conditions (also when looking at the individual participant data in supplementary figures S2.5, S2.6) despite the fact that the targets differed in their characteristics. The alarm was louder than the beep, it always came always from the same direction, and was the only sound coming from that direction. The beep, originated from different directions and was embedded into the other sound streams. The behavioral results show that the alarm was easier to detect than the beep. The reaction time for the detected sounds was not significantly different. This shows that a sound which is hard to detect, i.e., acoustically not salient, can elicit a clear attention response if it is considered task-relevant.

Processing of irrelevant stimuli Regarding the irrelevant sounds, we did not find a difference between conditions in the N1. We expected that attention to the beep (i.e., the wide condition) would draw attention to the whole soundscape and in turn also lead to a stronger processing of the irrelevant sounds. This manipulation was apparently not strong enough to produce a difference in the N1 component.

Nevertheless, we can draw a conclusion from the observed ERP morphology. The alarm and irrelevant sound elicited an N1 with a clear peak and strong deflection ($\sim$7-8 μV), while the beep elicited an N1 that was smaller ($\sim$2-3 μV) and smeared. We interpret the clear peak of the alarm and irrelevant sounds as an indication that these sounds were easily detectable (i.e., acoustically salient) compared to the beep as the N1 is sensitive to sound intensity (Näätänen, 1982; Näätänen & Picton, 1987). Furthermore, early auditory responses indicate awareness of a stimulus (Schlossmacher, Dellert, Bruchmann, & Straube, 2021). Thus, the clear peak of the alarm and irrelevant sounds might indicate that these sounds showed a different early processing compared to the beep.

Processing of the soundscape as a whole We found reliable TRFs in response to the complex soundscape (including language and non-language stimuli) in this complex task, with three time-windows which significantly differed from zero. These time-windows have repeatedly been reported for speech and music stimuli (Hausfeld, Riecke, Valente, & Formisano, 2018; Horton, D'Zmura, & Srinivasan, 2013; O'Sullivan et al., 2015), however not for other complex soundscapes.

We further expected a difference in processing of the whole soundscape between the two conditions. We used the beep in the wide condition to implicitly direct the participants attention toward the whole soundscape. In the narrow condition most of the soundscape could be ignored. We found a significant but small increase of processing in the wide condition after controlling that the effect was not due to the targets. Interestingly, the difference appeared in the last time-window. When tracking the response to an attended and continuous speech stream, an enhanced responses in late time-windows is observed compared to an ignored stream (Holtze et al., 2021; Horton et al., 2013; Jaeger et al., 2020; Kong et al., 2014; Mirkovic, Debener, Schmidt, Jaeger, & Neher, 2019; Petersen et al., 2017). As we expected that participants attend to the whole sound scape more in the wide compared to narrow condition, it is plausible that we observed a difference in the last time-window.

There are several reasons why the observed difference was small. On the one hand, participants did not attend to the whole soundscape much more in the wide than in the narrow condition. This would also explain the low hit rate for the beep. On the other hand, there was no incentive to ignore the soundscape in the narrow condition, which might have increased the response to the soundscape in the narrow condition. We conclude that our results are an indication that differences in the processing of the whole soundscape are found in late time-windows.

Random effects of the models We further investigated the random effect structure for a better understanding of the variance that contributed to our models (Lorah, 2018; Volpert-Esmond et al., 2021). Interestingly, the between-person ICC of the response to the alarm was twice as large compared to the beep. This indicates that in naturalistic soundscapes, reliable sounds (such as the alarm which was presented from the same direction with the same sound intensity) produce a more reliable trial-level response than unreliable sounds (such as the beep which was presented from different directions and with different sound intensities).

The low between-channel variance indicates that we used channels that were related to the investigated components (i.e., N1 and P3; Volpert-Esmond, Merkle, Levsen, Ito, & Bartholow, 2018). Furthermore, the selected channels had a close proximity. For the beep we even had to exclude channel as a factor, as we ran into singularity issues.

Translation to the operating room We designed our study to contain several factors that characterize the working environment in an operating room, i.e., multiple sound streams from different locations with relevant and irrelevant sounds, speech and non-speech sounds, and a visual-motor task. Our results demonstrate that it is feasible to study auditory attention in such a complex scenario. We observed a clear N1 peak for sounds that were acoustically salient, a P3b for relevant sounds, and a TRF in response to the whole soundscape. Thereby, our study is a step toward studying auditory responses in the operating room using mobile EEG.

We manipulated the attentional focus of participants who were naive to the soundscape of the operating room, thus the soundscape was rather arbitrary to them. This way, our results are generalizable to other scenarios with similar complex soundscapes. A limitation of this approach is that medical staff might react differently to the soundscape, because they are regularly exposed to it. Eventually, we want to know how the individual perceives the soundscape in the operating room and when sounds become a burden. Therefore, one must be aware of the challenges of studying sound processing in the operating room: First, the soundscape of an operating room is an uncontrolled setting in which the presentation of sounds is ethically not viable. Therefore, it is necessary to relate the natural soundscape to the EEG recording. We showed that meaningful EEG responses can be measured in a complex soundscape. In a following step, we suggest using smartphone-based technology that enables the simultaneous recording of EEG and audio features in a data protected way (Blum, Hölle, Bleichner, & Debener, 2021; Hölle, Blum, Kissner, Debener, & Bleichner, 2022), and applying it to the operating room. This way, responses to naturally occurring sounds can be measured. Second, surgery staff are exposed to the soundscape for several hours per day. Investigating changes in sound processing over the day requires long-term recordings. Here, one could use minimal and unobtrusive EEG set-ups, like the cEEGrid (Debener, Emkes, De Vos, & Bleichner, 2015), that can be used to study EEG responses to auditory stimuli (Holtze, Rosenkranz, Jaeger, Debener, & Mirkovic, 2022; Meiser, Tadel, Debener, & Bleichner, 2020) over more than 6 h (Hölle et al., 2021). Lastly, the cognitive load (e.g., working memory and attentional capacities) varies during a surgery over time and between staff members, which likely affects auditory processing. van Harten et al. (2021) observed that surgery staff with high workload are more often distracted by irrelevant sounds than surgery staff with low workload. However, this relationship is simplified as high load can also reduce the processing of irrelevant sounds (Brockhoff et al., 2022). Studying the relationship between load and auditory processing in the operating room is therefore necessary to understand the effect that sounds have on surgery staff.

Conclusion We showed that ERPs, as well as TRFs, are useful tools to study different aspects of sound perception in complex sound environments. To balance between high control over stimuli and the uncontrolled operating room we developed a laboratory experiment with a naturalistic soundscape. In this scenario, ERPs are robust to detect attention responses to specific sounds while TRFs can measure responses to an uncontrolled soundscape. Our results demonstrate that we can use mobile EEG in a complex acoustic-visual-motor task to study auditory perception and are therefore an important step toward understanding auditory attention in uncontrolled settings.

2.6 Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Zenodo (https://zenodo.org/record/7147701).

2.7 Ethics statement

The studies involving human participants were reviewed and approved by Medizinische Ethikkommission, University of Oldenburg, Oldenburg. The patients/participants provided their written informed consent to participate in this study.

2.8 Funding

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the Emmy-Noether program–BL 1591/1-1–Project ID 411333557 and by the Forschungspool Funding of the Oldenburg School of Medicine and Health Science.

2.9 Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

2.10 Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

2.11 Supplementary material

Deviations from our preregistration

General deviations:

1. We originally referred to the beep as the odd tone.

Deviations from planned preprocessing:

1. We did not specify the ICLabel threshold of .9 to reject components.
2. Regarding the P3, we initially searched for a peak between 250 and 400 ms. However, for some participants this resulted in finding the the P2 and not the P3 time-window. Therefore, we searched for a peak between 300 and 400 ms.
3. Regarding the data epochs for the TRF, we did not consider that the total length might significantly vary between conditions. Thus, epoch exclusion until the conditions have a similar length was added as a preprocessing step.

Deviations from planned analyses:

1. We did not specify during the preregistration how we deal with models that run into singularity issues, but trimming random factors with zero variance is recommended (Volpert-Esmond et al., 2021). Thus, we decided to drop the random factor "channel" for the ERP analyses of the beep.
2. As proposed in the preregistration, we planed post-hoc power analyses. However, using trial-level data our computational power was not sufficient to calculate a power analysis (i.e., it would have taken several weeks to conduct such analysis). Besides, we noticed that the use of post-hoc power analysis might be misleading, as it diverts from the true power to detect a significant effect (Zhang et al., 2019).

Supplementary Method

Tetris Game Design

The Tetris space consisted of 15 layers and each layer covered a 5x5 area. The layers and area formed a grid with equally sized cubes, thus forming a space for 15x5x5 cubes (see Figure 2.1). A block was formed from several cubes, resulting in blocks of ten different shapes (see Figure 2.1). A block was moved with the arrow keys or rotated cloackwise with the A,S,and D keys. The blocks dropped at a fixed speed until they were placed at the bottom of the space or on top of other block(s). While a block dropped, it was transparent and only its boarders were visible. If a block was placed, it was colored depending on the layer in which it was placed. Each layer has its own color. Thus, if a block covered more than one layer, its cubes were colored differently. To the right of the Tetris space, the current score, the shape of the next block, and the color code of the layers was shown.

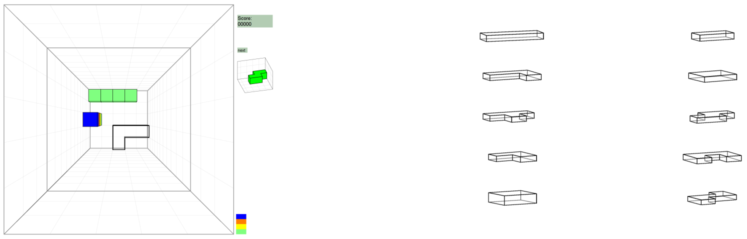


Figure S 2.1: Tetris box (left) and blocks (right)

Within the game, there were three ways how points were granted or subtracted:

First, if a layer was completely covered with cubes, it was deleted and 100 points were granted. If the blocks stacked too high, i.e., if at least one cube exceeded the 13th layer, blocks in all layers were deleted and 100 points subtracted. Thereby, participants had enough time (i.e., 3 layers) to navigate the blocks even if blocks were stacked high. Furthermore, by deleting all blocks, continuity of the game was guaranteed.

Second, participants received instructions to press the space bar when hearing the targets, i.e., the alarm or beep in the narrow or wide condition, respectively. Correct presses granted 50 points, missing a target subtracted 50 points.

Third, following the task-relevant speech granted 100 points while not following it subtracted 100 points. The task-relevant speech gave instructions to place the block in the upper or lower and left or right corner of the 3D space. A corner was defined as a 2x2 area in any layer. If at least one cube of a block was placed in the correct area, the points were granted.

Auditory Stimulus Material

Three types of auditory stimuli were integrated: a continuous background sound, distinct sound events, and speech. The background sound was adapted from an open-access hospital recording available on youtube [9]. The adaptation was manually done in Audacity® and included the removal of identifiable speech segments and creating chunks of the audio which could be rearranged and duplicated. This way, two versions of the background sound were created which were counterbalanced across conditions.

Regarding the distinct sound events, four sounds were included. Two of them were hospital monitor sounds which served as the irrelevant sounds[10] and one a hospital alarm sound[11]. As the youtube soundfiles contained multiple presentations, one sample of each sound lasting approximately 200 ms was manually extracted with Audacity®. The irrelevant sounds were presented equally often in the experiment. The fourth sound event, the beep, had a frequency of 800 Hz, was about 60 ms long, and was generated in MATLAB.

[9] https://www.youtube.com/watch?v=qR9YzVqO9Zg

[10] https://www.youtube.com/watch?v=4NXe9pwEgN4&list=OLAK5uy_m0wcbXgo3PYKOigdAPPPf0s9y2F7ZXtp0&index=127

[11] https://www.youtube.com/watch?v=95siaTtQR-c&list=OLAK5uy_m0wcbXgo3PYKOigdAPPPf0s9y2F7ZXtp0&index=128

Two types of speech stimuli were presented. The first type was task-relevant speech. It was generated using a text-to-speech program[12]. The speech instructed participants to place the next block in one of the four corners, thus four instructions were created. Each of the four instructions were presented equally often and the same instruction was never presented subsequently. The second type of speech was task-irrelevant speech. It contained snippets from a German podcast conversation between a male and female speaker. The snippets were manually extracted using Audacity® and presented in a predefined order. Each snippet lasted on average 3.5 (+- 1.5) seconds. Two podcasts were used [13] which were counterbalanced across conditions.

After the initial manual extraction of the sounds, they were further processed in MATLAB. Therefore, each stimulus was matched to the average root-mean-square value of all sounds. The sounds were spatially seperated using the Head Related Impulse function (Kayser et al., 2009), except for the background sound. As the loudness of the sounds varied, they were multiplied by individual factors (i.e., gain) before the spatial separation algorithm was applied. As the beep is presented together with other sounds, five versions of the beep were created. Each version was processed with the same parameters as the respective stimulus. For the beep in the background, no spatial separation was performed. Table S2.1 provides a detailed overview of the applied parameter for each stimulus.

Stimulus	Location	Listener Position	Speaker Position	Gain
Irrelevant Sounds	Cafeteria	2	A	10
Alarm	Cafeteria	2	E	10
Relevant Speech	Cafeteria	2	D	25
Irrelevant Speech	Cafeteria	2	B	6

Table S 2.1: HRIR processing parameters of auditory stimuli.

Stimulus Presentation

At the start of the game, the 3D-space was empty. A game started after a countdown counted from three to one. A Tetris block always started in the middle of the upper most layer, i.e. the 15th layer. The shape of a block was randomly chosen from one of the ten shapes.

Within the game, auditory stimuli were presented using Psychtoolbox 3 (Kleiner et al., 2007). For each presented stimulus a trigger marker was generated using a Lab Streaming Layer based software [14]. The background sound was presented throughout the game. Each of the four relevant speech instructions were presented 12 times. Each snippet of the irrelevant instructions was presented once in a predefined order. In total 48 irrelevant

[12]www.notevibes.com(Notevibes, 2021)

[13]https://www.ndr.de/wellenord/Allein-unter-Moerdern-Sabine-Thiesler,kunstmichmal166.html and https://www.ndr.de/wellenord/Beruehren-verboten-Nicht-bei-Julia-Brunner,kunstmichmal150.html

[14]www.github.com/labstreaminglayer/liblsl-Matlab

speech snippets were presented. The two irrelevant sounds were presented 24 times each. The alarm and beep were presented 48 times each. However, the beep could occur together with other stimuli. At the start of the game the beep was randomly positioned in each stimulus. The onsets between the irrelevant sounds/alarm and beep were close, potentially influencing the ERP analysis. Therefore, for each time a sound was presented together with a beep, the sound and beep were again added to the pool of stimuli. This resulted in a variation of sound presentations across conditions and participants. For the ERP analyses only the 48 trials without other interfering sounds were used.

Auditory stimuli were presented randomly depending on the layer of a block: The first auditory stimulus during a block occurred between layer 12 to 6, the second stimulus three to six layers after the first one, and so on. However, there were some restrictions. Task-relevant speech could occur consecutively, but the same speech instruction (i.e., placing the block in the same corner) could not occur consecutively to ensure that the blocks do not stack too high. The beeps did not occur during the first five stimuli. The position of the beep within another stimulus was randomly defined at the beginning of each condition. Thus, it was different for each participant and condition, but fixed for a stimulus. Note, that the beep could occur at a different position for each of the four instructions, two irrelevant sounds, and 48 task-irrelevant speech snippets. The participants were not aware of the presentation frequency or order. They were only informed about the approximate length of a game.

Training

To get acquainted to the game, participants received written instruction and performed four training games. During the first training, block rotation and placement was trained without any auditory stimuli and lasted ten minutes. During the first half, blocks dropped at a low speed. During the second half, blocks dropped at normal speed.

The second training game introduced the background sound and task-relevant speech, as it was relevant for both conditions, and lasted approximately one minute. Here, participants received visual feedback whether they correctly followed the instructions.

The third and fourth training games included all stimuli and were condition specific and therefore, played before the respective condition. They lasted approximately two minutes each. In the narrow condition training, participants received visual feedback whether they correctly detected the alarm. Prior to this training game the alarm was presented to them. In the wide condition training, participants received visual feedback whether they correctly detected the beep. Prior to this training game the beep was presented to them once alone and once included in the other stimuli.

A game ended after all stimuli were presented. For the games of each condition this was the case after approximately 18 minutes. For the speech training, five instructions were presented. For the condition specific training, each stimulus was presented twice.

Channel Layout

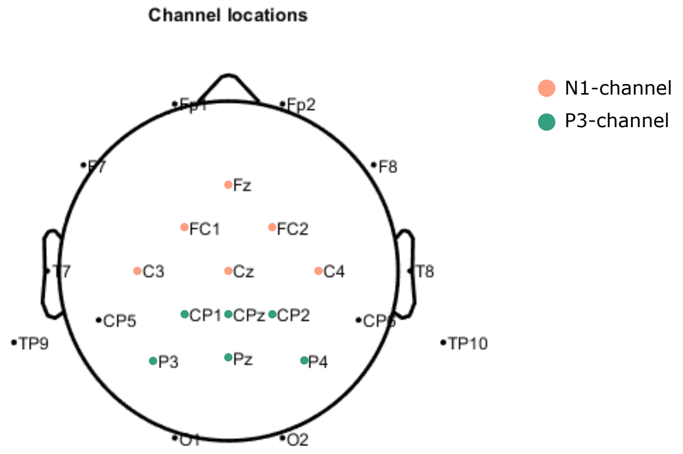


Figure S 2.2: A mobile 24-channel setup was used for this study. The colours orange and green indicate the channels that were used for the N1 and P3 ERP analysis, respectively.

Tetris performance

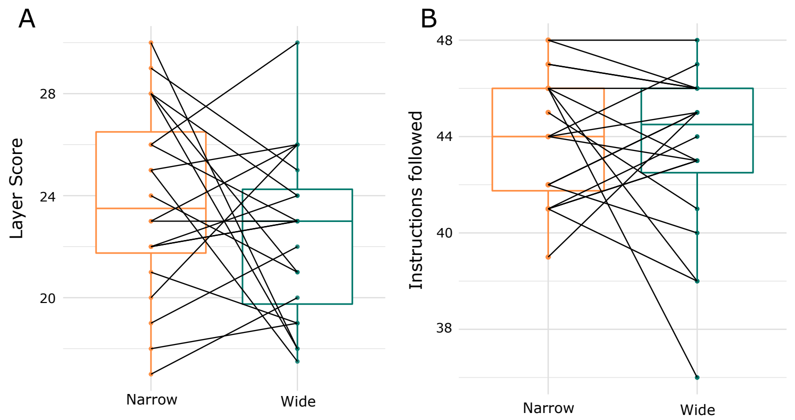


Figure S 2.3: **A)** Number of completed layers during the game. A layer was completed (and points added) if it was fully covered with Tetris blocks. **B)** Number of instructions (i.e., where to place the next blook) that were followed. Each line represent one subject. Both measures of task performance did not significantly differ between conditions.

Random effect model results

Stimulus	Group	Variance	Std. Dev
Alarm	Participant	12.08	3.476
	Channel	0.11	0.325
	Residual	86.71	9.312
Beep	Participant	6.002	2.450
	Channel	0	0
	Residual	94.028	9.697
Irrelevant sounds	Participant	13.44	3.6661
	Channel	0.162	0.4026
	Residual	76.9825	8.7740
TRF III	Participant	3.019	1.737
	Residual	1.099	1.049

Table S 2.2: Random effect results of the intercept model for the alarm, beep, irrelevant sounds, and the third TRF time-window

Long time-window of target ERPs

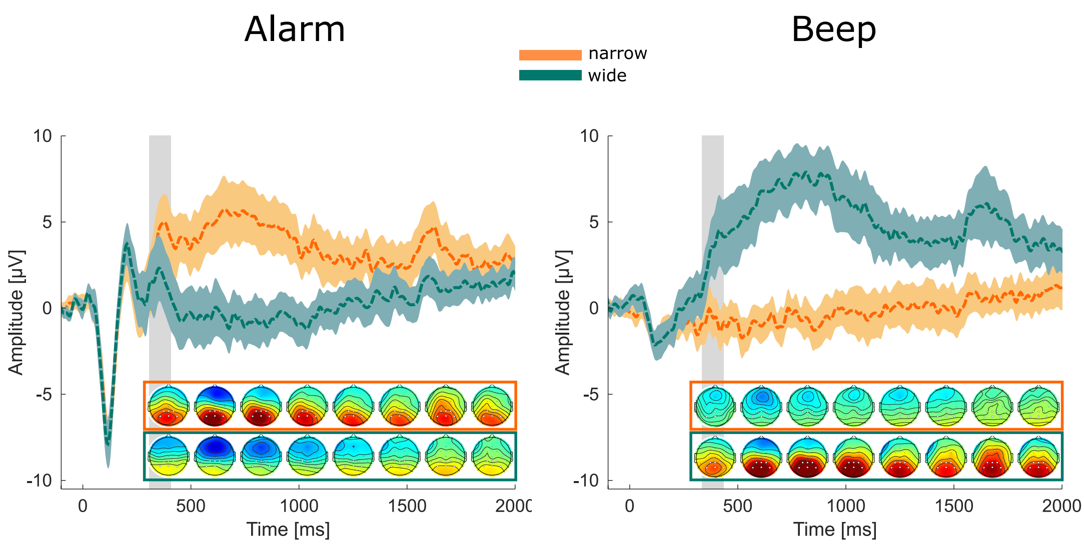


Figure S 2.4: The ERPs are the same as in figure 2.3. Topographies show time-windows from 300 to 1900 ms over time in steps of 200 ms.

Individual participant ERPs

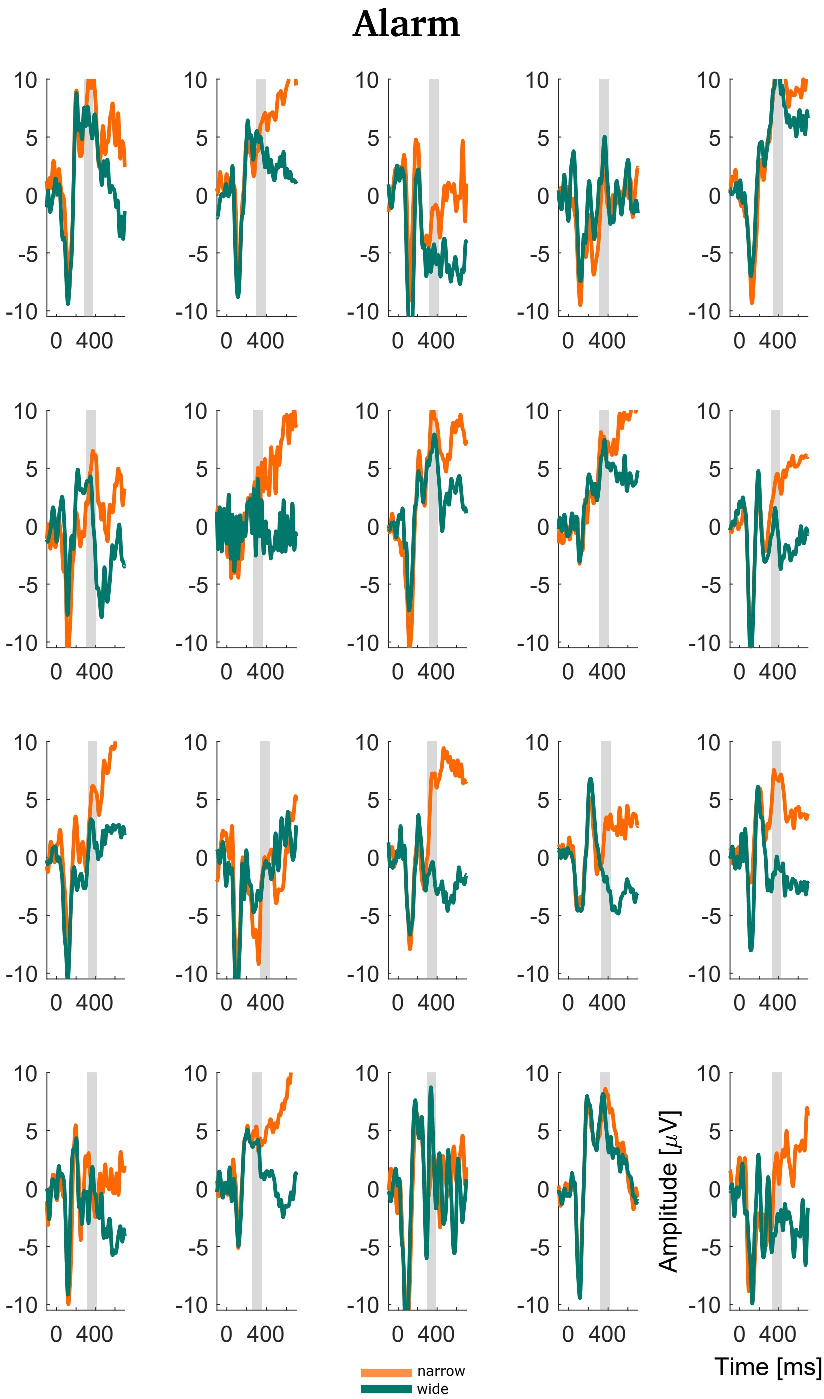


Figure S 2.5: Individual participant data in response to the alarm averaged over selected P3-channel and trials. Gray area marks the individual time-window used for the statistical comparison.

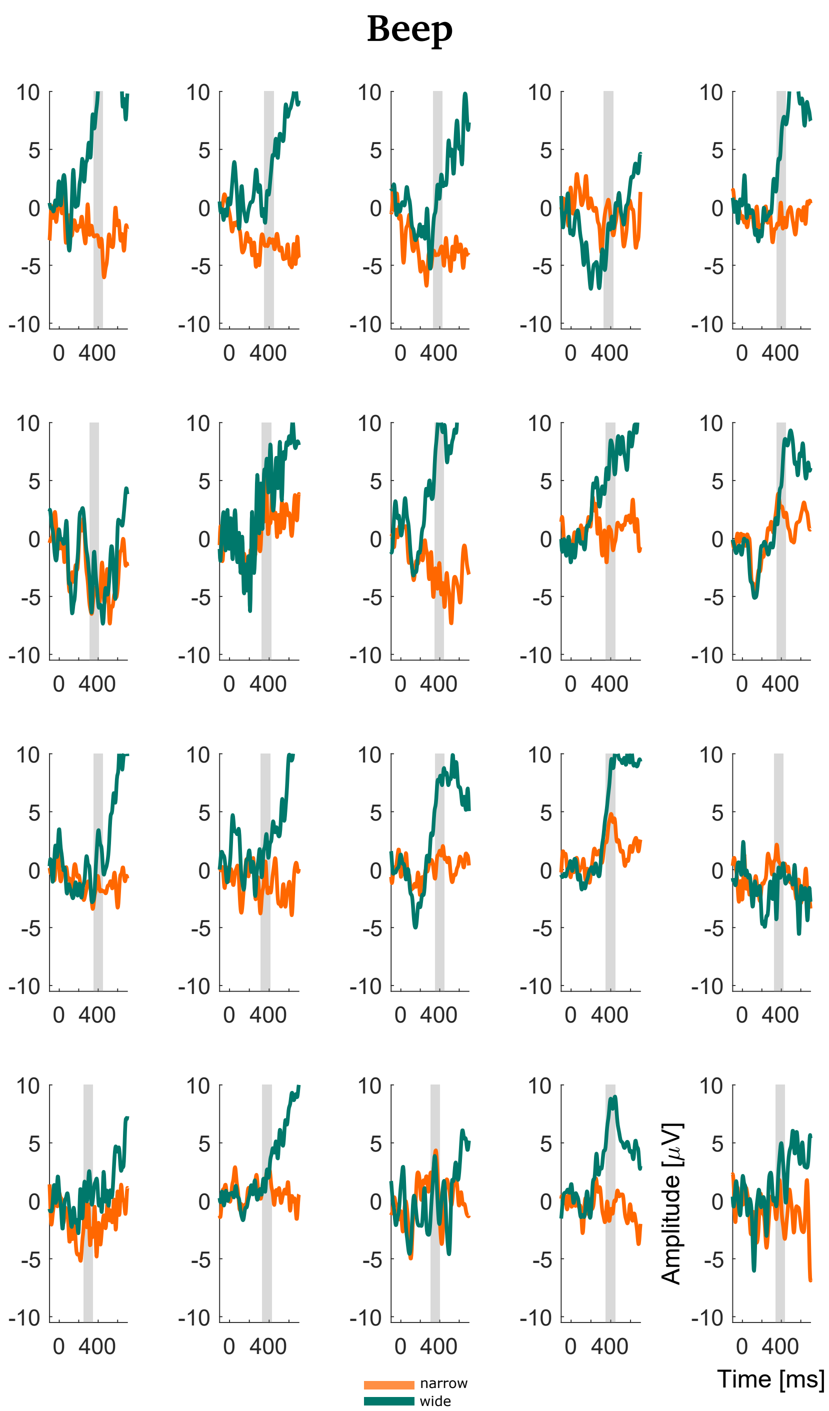


Figure S 2.6: Individual participant data in response to the beep averaged over selected P3-channel and trials. Gray area marks the individual time-window used for the statistical comparison.

Irrelevant Sound

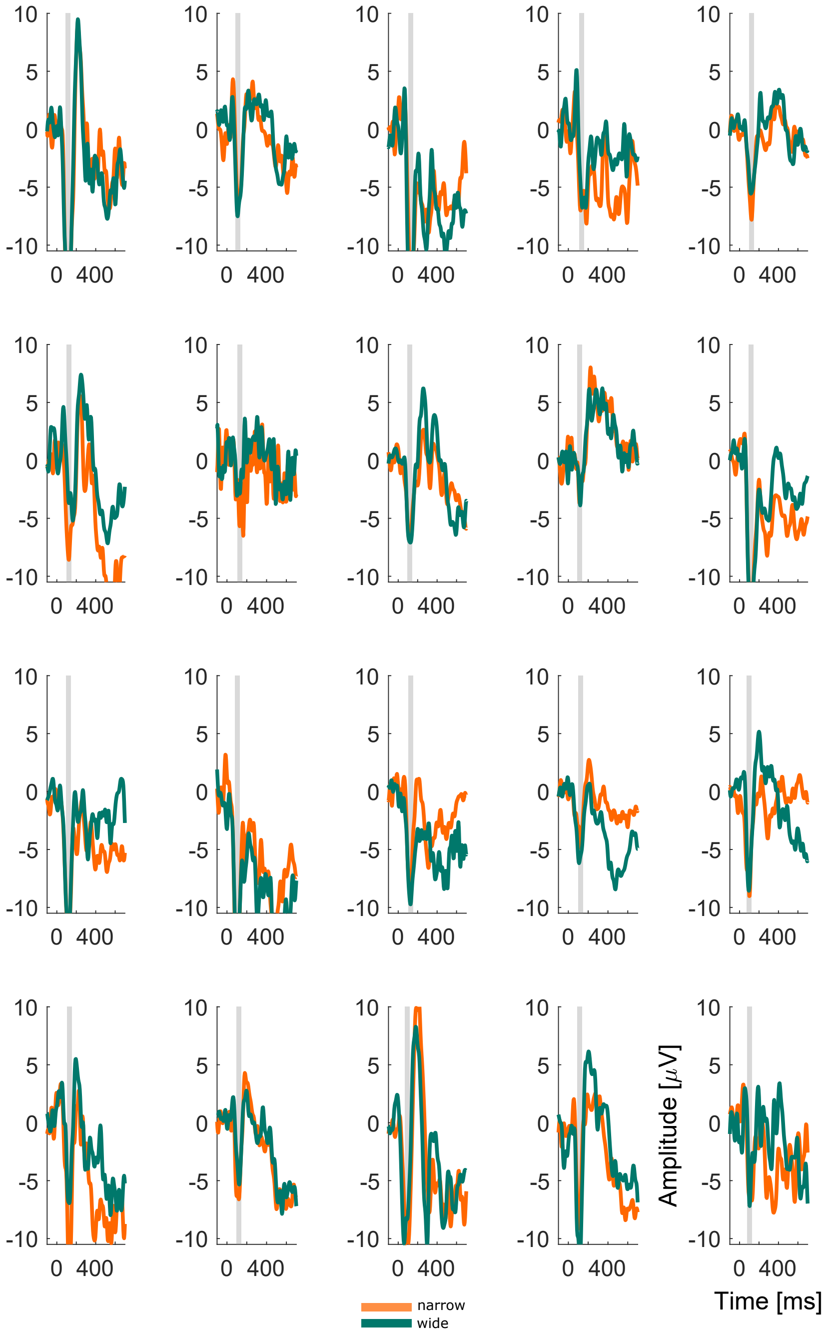


Figure S 2.7: Individual participant data in response to the irrelevant sounds averaged over selected N1-channel and trials. Gray area marks the individual time-window used for the statistical comparison.

Individual participant TRFs

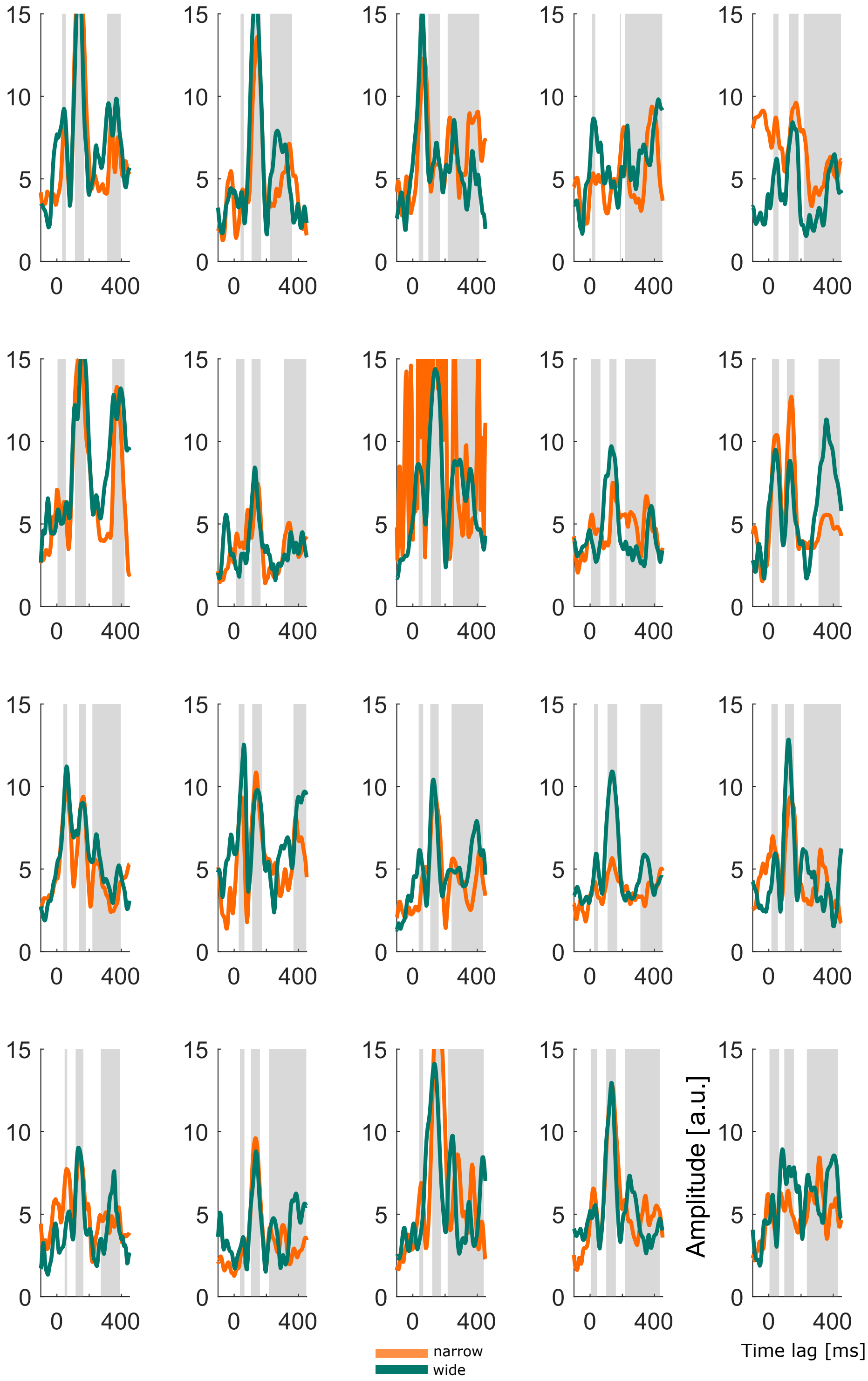


Figure S 2.8: Globald field power of the TRF for each participant in response to the whole soundscape. Gray areas mark the individual time-windows used for the statistical comparison.

3 Study II - Using mobile EEG to study auditory work strain during simulated surgical procedures.

This article has been published as:

Rosenkranz, M., Haupt, T., Jaeger, M., Uslar, V.N., & Bleichner, M.G. (2024). Using mobile EEG to study auditory work strain during simulated surgical procedures. *Sci Rep*, 3. https://doi.org/10.1038/s41598-024-74946-9

Author contributions: M.R. and M.B. conceptualized the experiment. M.R. performed the data acquisition and analyzed the data, which was assisted by T.H., M.J., and M.B.. M.R. wrote the manuscript to which T.H., M.J., V.U., and MB contributed with critical revisions. All authors reviewed the manuscript.

3.1 Abstract

Surgical personnel face various stressors in the workplace, including environmental sounds. Mobile electroencephalography (EEG) offers a promising approach for objectively measuring how individuals perceive sounds. Because surgical performance does not necessarily decrease with higher levels of distraction, EEG could help guide noise reduction strategies that are independent of performance measures. In this study, we utilized mobile EEG to explore how a realistic soundscape is perceived during simulated laparoscopic surgery. To examine the varying demands placed on personnel in different situations, we manipulated the cognitive demand during the surgical task, using a memory task. To assess responses to the soundscape, we calculated event-related potentials for distinct sound events and temporal response functions for the ongoing soundscape. Although participants reported varying degrees of demand under different conditions, no significant effects were observed on surgical task performance or EEG parameters. However, changes in surgical task performance and EEG parameters over time were noted, while subjective results remained consistent over time. These findings highlight the importance of using multiple measures to fully understand the complex relationship between sound processing and cognitive demand. Furthermore, in the context of combined EEG and audio recordings in real-life scenarios, a sparse representation of the soundscape has the advantage that it can be recorded in a data-protected way compared to more detailed representations. However, it is unclear whether information get lost with sparse representations. Our results indicate that sparse

and detailed representations are equally effective in eliciting neural responses. Overall, this study marks a significant step towards objectively investigating sound processing in applied settings.

3.2 Introduction

Surgical personnel often experience high levels of stress, which can lead to severe health problems such as burnout (Etheridge et al., 2023) or hypertension (Marrelli et al., 2014; Rieger et al., 2014). One cause of stress is distractions due to the environment of the operating room (OR) (Arora et al., 2010; Kern et al., 2019). The soundscape in the OR is highly complex (Baltin et al., 2020), comprising of sounds that are crucial to the surgery (e.g, communication, tool usage), as well as sounds that are not crucial to the surgery (e.g., instrument clattering, phone ringing) and could be minimized to improve the work environment (Engelmann et al., 2014). The accumulation of the different sounds sources, leads to high sound levels which are often perceived as distracting and stressful (Gülşen et al., 2021; Healey et al., 2007; Keller et al., 2018; Padmakumar et al., 2017; Tsiou et al., 2008; van Harten et al., 2021). To guide interventions that aim at reducing stress induced by auditory distractions it is important to understand and measure how ongoing sounds in the OR affect the personnel.

An objective evaluation of the subjectively experienced burden of the OR soundscape is challenging. Previous studies that have focused on the effect of sound on performance (i.e., surgery task performance) have found that only under extreme and unnatural conditions, such as dichotic listening to two pieces of music (Conrad et al., 2010, 2012), but not under more natural conditions, such as a single stream of OR sounds (Han et al., 2022; Hodge & Thompson, 1990; Moorthy, Munz, Dosis, et al., 2003; Moorthy, Munz, Sarker, & Darzi, 2003; Moorthy et al., 2004; Suh et al., 2010), more mistakes in the surgery task were observable. However, medical personnel may strive to perform at high levels under adverse working conditions because mistakes can have serious consequences for the patient. Thus, the work strain surgeons experience can not necessarily be inferred from the surgery task performance. Therefore, additional measures are required to objectively assess the auditory strain in the OR.

To measure sound perception in the OR objectively, electroencephalography (EEG) is a promising method. Given its temporal resolution, it is particularly useful for measuring responses that are time-locked to sounds. By analyzing event-related potentials (ERPs) the neural response to individual sound events can be examined, revealing perceptual and cognitive processes such as whether a sound is considered task-relevant (Polich, 2007). Another approach involves the use of temporal response functions (TRFs) to study responses to continuous sound streams (Crosse et al., 2016), which expands the use of EEG to naturalistic soundscapes. ERPs and TRFs can be measured outside of the laboratory while individuals are freely moving (Debener et al., 2012; Straetmans et al., 2021) and to complex soundscapes while an audio-visual-motor task is being performed

(Rosenkranz, Cetin, Uslar, & Bleichner, 2023). The development of head-mounted EEG systems and wireless data transmission allows for unrestricted mobility for the wearer, making it applicable in various work environments (Wascher et al., 2021) and enabling measurements over extended periods of time (Hölle et al., 2021).

How we process distractors depends on contextual factors, including the demands of the task(Lavie, 2005), and is reflected in the neural responses measured by EEG (Brockhoff et al., 2022). Thus, EEG signals can provide valuable insights into the cognitive demand experienced by an individual, independent of their task performance. This makes EEG a potentially useful tool for continuously monitoring a surgeon's perceived demand in real time. Such monitoring could guide the implementation of sound interventions in the OR by addressing not only the acoustic environment (e.g., loudness Engelmann et al., 2014), but also the individual's perceived demand. Furthermore, cognitive demand influences the ability of a person to shield themselves from auditory distractions (Sörqvist et al., 2016). Personnel in the OR report to be more vulnerable to distractions from the soundscape during periods of high compared to low demand (van Harten et al., 2021). It is reasonable to expect that both neural and subjective responses to the OR soundscape will differ based on the cognitive demands of the task being performed. By varying the overall cognitive demand, we can study how it impacts perception and processing of soundscapes, which in turn helps us to understand auditory distraction.

We investigated demand-dependent changes in EEG responses to the soundscape, as well as subjective and behavioural measures. We focused on ERP and TRF time-windows typically found in response to sounds, namely the N1, P2, and N2 time-window (Horton et al., 2013; Kong et al., 2014; Picton & Hillyard, 1974). Previous studies reported mixed findings regarding the effect of varying demand on these time-windows (Brockhoff et al., 2022). Therefore, we investigated the effect of varying demand on each time-window.

Surgeons are frequently required to perform complex procedures for extended periods. This naturally leads to fluctuations in cognitive demand and how the environment is perceived. For example, sounds that were once easily ignored can become sources of distraction, or vice versa (van Harten et al., 2021). This change in sensory processing, compounded by varying demand, underscores the critical need to consider the temporal dynamics during the performance of surgical tasks. Therefore, we explored changes over time for the ERP and TRF time-windows in response to the soundscape, as well as for the subjective and behavioral measures.

At last, there are practical considerations when opting to use TRFs in an applied context. Since the derivation of TRFs require information of the soundscape, sound recordings have to adhere to privacy concerns. To test whether representations of the soundscape void of personal information, produce similar results as rich representations we derived sound onset marker (Hölle et al., 2022) in addition to the commonly used acoustic envelope (Crosse et al., 2016). Sound onsets only indicate sound occurrence, providing a data protected way of sound recording. Previous research has demonstrated that onsets and envelopes yield comparable results for computing TRFs (Drennan & Lalor, 2019; Haupt, Rosenkranz, & Bleichner, 2024). To replicate this finding and strengthen the applicability of TRFs in real-life settings we compared the TRFs computed from the onsets and the envelopes of the soundscape.

The objective of this research was twofold. Firstly, it aimed to examine the influence of a realistic task under varying levels of cognitive demand on the individual processing of task-irrelevant auditory stimuli, reflecting an OR soundscape. This investigation integrated both subjective assessments (i.e., self-reports) and objective metrics (i.e., neural responses to irrelevant sounds and behavioral performance) to provide a comprehensive understanding of sound and distractor processing in work environments. Secondly, the study investigated the processing of continuous versus discrete acoustic features, offering insights into auditory responses to different aspects of the soundscape.

3.3 Methods

This study involving human participants was approved by the Medizinische Ethikkommission, Carl von Ossietzky Universität Oldenburg, Oldenburg (2021-031) and performed in accordance with the Declaration of Helsinki. The participants provided their written informed consent to participate in this study. This study was preregistered after data collection https://doi.org/10.17605/OSF.IO/AE3UY. All changes to the preregistration can be found in the supplementary material.

3.3.1 Participants

23 medical students were recruited through an online announcement on the university board and word-of-mouth (age range: 19-24; 16 female). One participant was excluded due to data loss during recording resulting in 22 included participants. We based the sample size on our previous studies in which we measured reliable ERPs and TRFs during a complex and dynamic task (Rosenkranz et al., 2023). Four participants had previous experience with a laparoscopic simulator and 17 attended at least once a surgery as an observer. All participants received monetary reimbursement. Eligibility criteria included: enrollment as a medical student and self-reported normal hearing, normal or corrected vision, no psychological or neurological condition, and right-handedness.

3.3.2 Paradigm

Figure 3.1 illustrates the setup and paradigm. To mimic an actual surgical environment, participants performed a laparoscopic surgery task while presented with the playback of a continuous OR soundscape. To vary the cognitive demand during the surgery task, we adapted a serial recall paradigm: Prior to the surgery task, participants had to remember two or eight letters, simulating low and high cognitive demand, respectively. They received the instruction to silently repeat those letters while performing the surgery task. After the surgery task, they should recall the letters. The memory task was selected for the following reasons. First, to simulate a varying general demand of surgeons that might not be necessarily related to the surgery. Second, the task was chosen as visual working memory abilities correlate with surgical task performance (Hedman, Klingberg, Enochsson, Kjellin, & Felländer-Tsai, 2007; Schlickum et al., 2011), i.e., we manipulated the demand on the surgery task by manipulating the demand on visual memory. Third, the memory task did not require an overt response and hence did not interrupt the surgical procedure. Finally, we could increase the distractive potential of the task-irrelevant soundscape, by presenting spoken letters which are commonly used as distractors in classic serial recall paradigms (Colle & Welsh, 1976; Jones & Morris, 1992). The spoken letters were presented in the low- and high-demand condition. Overall this should simulate situations in which the surgeon must effectively shield themselves from irrelevant and potentially distracting information. Participants did not receive any instruction on what to do with the soundscape, and were free to disregard it. After retrieval of the letters, participants rated their perceived demand during the task and received feedback regarding their memory performance. Each participant performed a total of 28 experimental blocks, 14 of each condition, in a randomized order, with the restriction that seven blocks of each condition were presented during the first and second half of the experiment. During the entire experiment, participants were standing either in front of the screen for the memory task (Samsung, SyncMaster P2470) or in front of the surgery simulator. The monitor and simulator were positioned in 90° to each other. Further details can be found below.

Surgery task For the surgery task, the LabSim® (Surgical Science, Sweden) simulator was used. This simulator is used to train hand-eye coordination for laparoscopic surgeries and provides realistic graphics and tactile feedback. The setup includes specialized instruments for both hands, a foot pedal for additional controls, and a screen that displays the virtual surgical instruments in action.

The specific task which the participant performed throughout the experiment is called "cutting". In this task, participants needed to grasp a vessel with the right-hand instrument, stretch it, and cut it with an electrical cutting device which was handled with the left hand. To activate the cutter, a foot paddle must be pressed. The vessel must then be placed in a small bag. The task ends when three parts of the vessel are removed.

Participants were informed that their performance would be evaluated based on the duration to complete the surgery task, mistakes, and tissue damage. Mistakes included rupturing the vessel (e.g., by exerting too much pull on the vessel) or dropping the cut vessel, thus a block could have zero to three mistakes. Touching the tissue with the instrument caused immediate visual feedback as the screen received a red shade. After each task, they received visual feedback about their overall performance.

Memory task A serial recall task was adapted to vary the cognitive demand during the surgery task. Although the serial recall task is a classic working memory task, we refrain

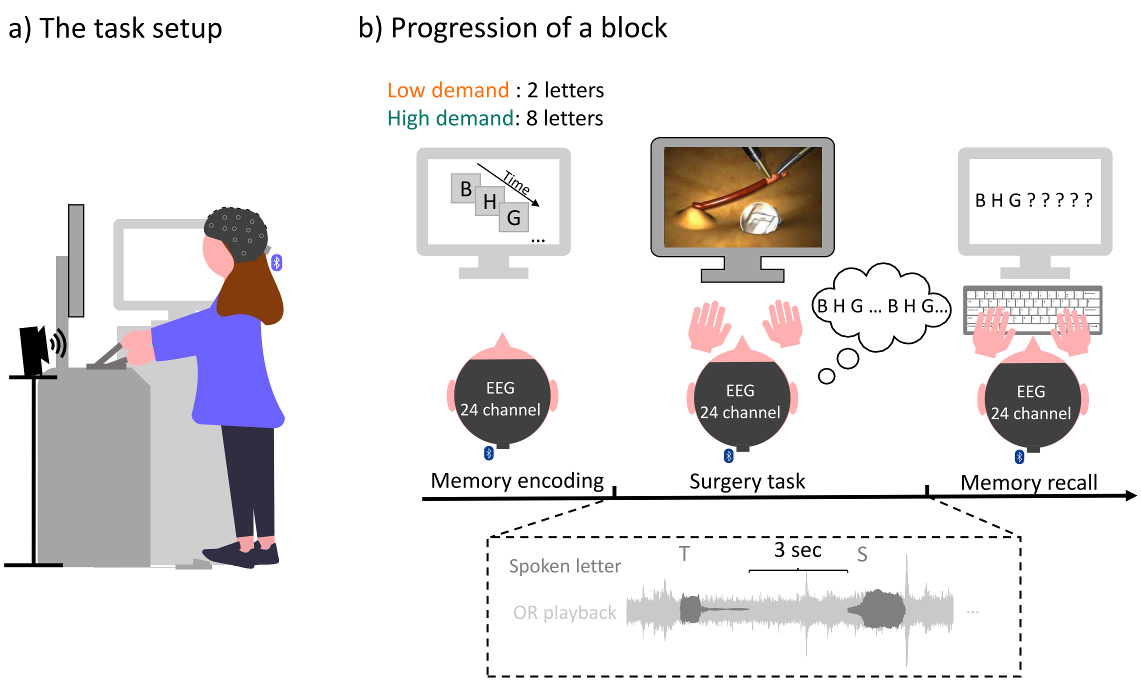


Figure 3.1: a) Participants were standing throughout the entire experiment either in front of the surgery simulator or the screen for the memory task which were positioned in 90○ to each other. The speaker were positioned at chest height and participants were equipped with a mobile EEG cap. b) At the start of the block participant were presented with either two (low-demand condition) or eight (high-demand condition) letters which they had to remember. During the surgery they were presented with a continuous realistic soundscape of an operation room and with overlayed spoken letters. They were instructed to silently repeat the memorized letters during the surgery task. After finishing the surgery task they should enter the memorized letters and fill out a workload questionnaire. An experimental block ended with feedback regarding their memory performance.

from calling it one, as the retention interval (i.e., the performance of the surgery task) was too long. Prior to the surgery task, participants were asked to remember the visually presented letters in the correct order. For the low-demand condition two letters and for the high-demand condition eight letters had to be remembered. The letters were randomly selected from a set of twelve letters (B,C,D,F,H,K,L,M,P,Q,S,T) without replacement. Letters were presented in black on a gray screen for 1000 ms each with an inter-stimulus-interval of 500 ms. After the surgery task, either two or eight question marks indicated that the letters should be entered using a keyboard. Participants could enter an "X" for letters they could not remember and could correct themselves.

Soundscape The OR playback was recorded using a field recorder which was positioned close to the surgery table during a visceral surgery at the University Hospital Oldenburg (Rennies et al., 2023). The recording contains a variety of sounds, such as ventilation noise, beeps from monitoring devices, instrument clatter, and instrument sounds. Intelligible speech was removed after the recording for privacy reasons, however, unintelligible muttering and non-vocal sounds such as coughing were preserved. The recording lasts approximately 1 hour. In the first block, the recording starts at a random time point and continuous chronologically for every subsequent block. If the recording reached the end, it started at the beginning. The soundscape started automatically after the letters of the memory task were presented. To prevent clicking noise when the audio starts the starting 500 milliseconds were faded in.

In addition to the realistic soundscape we presented a sequence of individual spoken letters in both conditions. The method of presentation of the spoken letters was consistent across both conditions: Four letters were drawn from the same set of letters as the memory task, but never coincided with the to-be-remembered letters of a block. To ensure that the letters were presented equally often within a block, the four letters were presented as consecutive groups but were randomized within a group. The same letter was never presented consecutively. The inter-trial-interval between letters was three seconds. Letters were generated using a text-to-speech program (Notevibes, accessed 2021) and spoken by the same female voice. The letters were clearly audible in the soundscape recording and had an on- and offset ramp of ten milliseconds.

All sounds (i.e., the recording and letters) were sampled at a rate of 48 kHz and presented to the participant using Psychtoolbox 3 for MATLAB(Kleiner et al., 2007) (v3.0.17), a t.amp E4-130 amplifier (Thomann GmbH, Burgebrach, Germany) and presented as a stereo signal using two iLoudMTM loudspeakers (IK Multimedia Production srl, Modena, Italy). The loudspeakers were vertically tilted upwards (20°) and located in front of the participant to the right and to the left of the LabSim® at chest height. The distance between loudspeakers was 0.5 m and the distance between loudspeaker and ear was 1.2 m. The sound pressure level was 45-55 dB SPL, measured at the place of the participants head.

Subjective workload assessment To assess the subjectively perceived demand during each block we included three workload related questions. For a good representation of our research question we chose two items from the NASA-TLX (Hart & Staveland, 1988), namely effort ("How hard did you have to work to accomplish your level of performance?") and frustration ("How insecure, discouraged, irritated, stressed, and annoyed were you?"). The effort item was chosen to investigate whether remembering eight compared to two words was perceived as more effortful, thereby linking it to our investigation of differences in demand. The frustration item was included to assess the emotional response to the high-demand task, particularly considering the potential stress induced to remember eight compared to two letters. Furthermore, one question from the SURG-TLX (Wilson et al., 2011) was chosen, namely distraction ("How distracting was the operating environment"). This item was selected to investigate how the perception of the soundscape interacts with task-related demand. Each question was answered on a visual analog scale ranging from 0 to 20 (Hart & Staveland, 1988; Wilson et al., 2011).

Training To familiarize themselves with the tasks, participants engaged in a series of practice blocks. During these sessions, the experimenter remained nearby, providing guidance and support to help participants navigate through the tasks. First, they performed the memory task twice, once with two letters and once with eight letters. Second, they performed a basic surgery task (i.e., instrument navigation) without the memory task or sounds to familiarize themselves with the LabSim®. Third, they performed the cutting task without the memory task or sounds. Lastly, participants performed two training blocks that were identical with the blocks of the main experiment, first for the low-demand condition and then for the high-demand condition. After the training, the experiment started and participants performed the experimental blocks on their own.

3.3.3 Data acquisition

Participants were asked to wash their hair on the day of the recording and to not use hair styling products. EEG data were recorded using a wireless 24-channel amplifier (SMARTING, mBrainTrain, Belgrade, Serbia) attached to the back of the EEG cap (EasyCap GmbH, Hersching, Germany) with Ag/AgCl passive electrodes at 10-20 layout positions (Fp1 Fp2 AFz Fz F3 F4 F7 F8 Cz C3 C4 T7 T8 CPz Pz M1 M2 P3 P4 P7 P8 POz O1 O2) with the reference and ground electrode at position FCz and Fpz, respectively. The data were recorded using a sampling rate of 500 Hz, and transmitted via Bluetooth from the amplifier to a Bluetooth dongle (BlueSoleil) that was plugged into a computer.
After fitting the cap, the skin was cleaned using 70% alcohol. Abrasive gel (Abralyt HiCl, Easycap GmbH, Germany) was used for reducing electrical impedance and ensuring high-quality signal. Impedances were kept below 20 kΩ at the beginning and again checked at the end of the recording using the SMARTING Streamer software (v3.4.3; mBrainTrain, Belgrade, Serbia).
ECG data were also recorded on a laptop but were not part of the current analyses.
Experimental markers (e.g., sound markers) were generated using the lab streaming layer library (Kothe et al., 2024) (v1.14). The ECG recording laptop and EEG recording computer were connected via Lan. To synchronize all data streams, EEG data, ECG data, and experiment marker were collected in the Lab Recorder software (v1.14, https://github.com/labstreaminglayer/App-LabRecorder) and saved as one .xdf file on the EEG recording computer.

3.3.4 Preprocessing

EEG The EEG data were analyzed using EEGLAB (Delorme & Makeig, 2004) (v2022.0) in MATLAB R2020b (The MathWorks, Natick, MA, United States). For each participant, the continuous data were filtered with Hamming windowed FIR filter using the EEGLAB default settings: (1) high-pass: passband edge = 0.5 Hz (filter order = 3300, transition bandwidth = 0.5 Hz, cutoff frequency (-6dB) = 0.25 Hz); (2) low-pass: passband edge = 30 Hz (filter order = 220, transition bandwidth = 7.5 Hz, cutoff frequency (-6dB) = 33.75 Hz). For further analysis, only the EEG data during sound presentation were included, i.e., data during memory encoding and retrieval were removed. To minimize artifacts from switching between the surgery task screen and the memory task screen we removed the first and last five seconds of EEG data during sound presentation. Bad EEG channels were automatically rejected using the EEGLAB function *clean_artifacts* from the *clean raw data* plugin (Mullen et al., 2015) following a procedure described by Klug & Gramann (2021)(Klug et al., 2022). The function was executed over ten iterations with the following parameters: ChannelCriterion (0.8); ChannelMaxBrokenTime (0.5). The remaining parameters were turned off. If a channel was rejected in at least 50% of iterations, it was removed. A maximum of 5 channels could be removed to ensure that enough data were available for channel reconstruction. On average 0.72 ($\pm$ 1.28) channels were rejected.

After bad channel removal, the data were cleaned from artifacts using infomax independent component analysis (ICA). For ICA, a copy of the preprocessed data were created and high-pass filtered (passband edge = 1 Hz, filter order = 825, transition bandwidth = 1 Hz, cutoff frequency (-6dB) = 0.5 Hz), and cut into consecutive epochs of one second. Improbable epochs with a global (all channels) or local (single channel) threshold exceeding 5 standard deviations were automatically rejected using the *jointprob* function. ICA decomposition was applied to the remaining epochs. The resulting components were back-projected on the original preprocessed, but uncleaned data. The components were then classified using the EEGLAB toolbox *ICLabel* (Pion-Tonachini et al., 2019) with the 'lite' classifier which is better at detecting muscle artifacts than the default classifier (Klug & Gramann, 2021). Components belonging to the categories eye blink and movement or muscle movement with 60% confidence were removed. Note, that the ICLabel classifier did not classify all components correctly because it was trained on stationary data with a larger electrode setup than ours. Therefore, we manually checked the components and made the following adjustments: We detected ICs located at the mastoids, probably from muscle movement (see Supplementary Figure S3.2 for an example). As the mastoids were used for re-referencing we manually removed these ICs. Furthermore, lateral eye movement also required manual removal in some cases. Afterwards, previously rejected channels were interpolated using spherical interpolation. Lastly, channels were re-referenced to the linked mastoids (M1/M2).

Audio For technical reasons, a constant delay between the marker that indicates a sound onset and the actual sound presentation was quantified beforehand. To correct for this constant delay the marker was shifted by 30 ms. Furthermore, the onsets of the auditory letters were also corrected. The letters were embedded in a constant sound stream which might lead to an energetic masking effect of the first few milliseconds of a letter. This affects when a letter becomes audible and therefore, the time when a brain response occurs. To obtain a better estimate when the participants could hear the letters, we used the OnsetDetector app (Hölle et al., 2022) implemented in MATLAB which determined the first energetic peak of the letters. The markers were shifted between 0 to 12.83 ms (Supplementary Figure S3.1).

In order to relate the ongoing soundscape to the ongoing neural response, acoustic features

were extracted. For this, we only used the OR playback (i.e., without the letters). From the OR playback we extracted and compared three feature vectors, namely the envelope of the raw OR playback, the envelope of the noise-reduced OR playback, and the onsets of the OR playback.
The envelope of the OR playback was extracted using the *mTRFenvelope* function (Crosse et al., 2016, 2021) with default inputs. To reduce the noise in the OR playback from ventilation and running machines, we used a Wiener filter implemented in MATLAB (Plapous, Marro, & Scalart, 2006; Scalart, 2023). For this, we first high-pass filtered the OR playback at 1 Hz (filter order = 1000, transition bandwidth = 0.5 Hz, cutoff frequency (-6dB) = 0.00004 Hz). We then estimated the power spectral density of the noise using the first second of the OR playback, as it was representative of the static noise in the OR playback. The noise estimate was then subtracted from the remaining signal. Afterwards, we extracted the envelope from the noise-reduced OR playback using the *mTRFenvelope* function.
Onsets were calculated using the OnsetDetector App (Hölle et al., 2022) implemented in MATLAB. As we aim to detect onsets in naturalistic settings the raw audio was used. The resulting feature vector contained zeros (i.e., no onset) and ones (i.e., onsets).

3.3.5 EEG analysis

We performed two types of analysis. An ERP analysis to study the event-related responses to the onset of the spoken letters, and a TRF based analysis to study the response to the ongoing OR playback.

ERP calculation ERP analysis was performed for the spoken letters. For each letter, epochs from -200 to 600 ms with respect to the stimulus onset were extracted. A baseline correction from -200 to 0 ms prior to stimulus onset was performed. Improbable epochs with a global (all channels) or local (single channel) threshold exceeding 3 standard deviations were automatically rejected using the *jointprob* function. We then computed the average response of each participant and block.

TRF calculation A forward modeling approach was used to compute a temporal response function (TRF) that characterizes the brain's temporal response to a feature vector representing the auditory stimulus. To calculate the TRF we used the mTRF toolbox (Crosse et al., 2016). For the TRF analyses, EEG data were multiplied by factor .0313 for scaling (as suggested in the provided code by Crosse et al. (2016)(Crosse et al., 2016)).
To evaluate which acoustic features best predicts the neural response we implemented a forward model based on individual EEG data using a 10-fold cross-validation approach. For this, we separated the blocks into 10 segments. With 28 blocks in total, each segment consisted of two to three successive blocks. We split the segments into training and testing data such that each segment was once test data and iterated through the following proce-

dure. For the training procedure we determined a shrinkage regularization parameter λ using the *mTRFcrossval* function with a time range from 0 to 450 ms time lag and a lambda range from 10^{-8} to 10^{8}. This resulted in a correlation value for each fold (i.e., number of TRF training blocks), lambda value, and channel. We averaged over folds and channels and used the lambda value that maximized the correlation for the subsequent TRF calculation. We then trained a forward model using the training blocks and the *mTRFtrain* function with time lags from 0 to 450 ms and the optimal lambda. The resulting model was used to predict the response of the test blocks using the *mTRFpredict* function. This resulted in a prediction value for each test block, channel, and iteration, which were averaged, leaving one prediction value per participant and acoustic feature.

Lastly, we computed a forward model for each block using *mTRFtrain* with time lags from -220 to 500 ms. For the individual optimal lambda value we chose the most frequently occurring one during the cross-validation procedure. As the scale of TRFs varied across participants, biasing statistical comparisons of amplitudes, we z-scored each participant's TRF weights across time-points, channel, and blocks.

GED analysis To evaluate amplitude differences of the ERP and TRF and avoid channel selection, we use generalized eigenvalue decomposition (GED) as a spatial filter following guidelines by Cohen (2022)(Cohen, 2022). In short, for each ERP and TRF time-window (i.e., N1/P2/N2) we computed a generic spatial filter across subjects that was then applied to the data. GED maximizes the contrast between a signal covariance matrix S and a reference covariance matrix R. For S, we computed N1, P2, and N2 time-windows and contrasted each time-window separately against the baseline period (i.e., -200 to 0 ms).

In detail, we first average ERPs and TRFs across blocks, resulting in one time-series for each participant. We then determined the N1/P2/N2 peak of each ERP and TRF. Regarding the ERPs, and TRFs calculated from the onsets we searched for the N1 peak between 80-150 ms, for the P2 peak between 150-250 ms, and for the N2 peak between 200-300 ms. The TRF calculated from the envelope showed earlier peaks, therefore we searched for the N1 peak between 50-120 ms, for the P2 peak between 120-220 ms, and for the N2 peak between 220-320 ms.

Second, we determined the channel with the largest amplitude for the N1/P2/N2 peak. Around the peak, we then calculated a time-window of $\pm$25 ms for the N1 and $\pm$50 ms for the P2/N2 peak.

Third, to obtain the corresponding GED filter weights we contrasted the ERP and TRF time-windows with the baseline period. Specifically, we mean-centered the data of the time-windows and baseline period and computed the covariance matrices for either time periods. For each participant, the covariance matrix of interest (S) was computed for the time-windows and the reference covariance matrix (R) was computed for the baseline period.

We cleaned the covariance matrix S and R by first computing the average covariance matrix $\bar{S}/\bar{R}$ across participants. We then computed for each participant the Euclidean distance between the covariance S and $\bar{S}$, and R and $\bar{R}$. Covariance matrices which deviated from $\bar{S}$

and $\bar{R}$ with more than 3 standard deviations were removed and $\bar{S}$ and $\bar{R}$ were computed again across the remaining covariance matrices. The $\bar{R}$ matrix was then regularized using a shrinkage regularization parameter of $\lambda = 0.01$.
The resulting spatial filter maximizes the contrast between the time-windows and baseline period. For each ERP and TRF time-window, this resulted in a number of eigenvectors that is equal to the number of channels. The eigenvector with the largest eigenvalue should best separate the baseline and time-window. However, this does not guarantee physiological plausibility. We therefore investigated the GED components and reported when not using the GED component with the strongest eigenvalue. As a result we received a GED component time-series for each participant, ERP/TRF time-window, and block, and for the TRF additionally for each acoustic feature. Additionally, we computed a forward model for each component to investigate the physiological interpretability of the filter.
As eigenvectors are sign uncertain, we set appropriate signs for the GED components (negative for N1, N2, and positive for P2). Lastly, we averaged the amplitude of the GED component across the ERP and TRF time-windows to extract one amplitude value per time-window, participant, and block and for the TRF also per acoustic feature.

3.3.6 Statistical analysis

All statistical analyses were performed in RStudio (v. 2021.09.0).
Prediction values of the acoustic feature were compared using a Wilcoxon signed rank test. We computed regression models for the subjective workload questions, behavioral responses, and GED component mean amplitudes for the ERP and TRF time-windows (i.e., N1/P2/N2). For these analyses, we started with a model including 'participant' as a random intercept and 'condition' as a fixed effect. 'Condition' contained two categories, i.e., low- and high-demand which were coded 0 and 1, respectively. We explored time-on-task effects (i.e., they were not part of the preregistration) by adding the block number as a continuous predictor 'time' in a second model. In the third and most complex model an interaction term between condition and time was added:

$$\hat{y} \sim condition + (1|participant) \tag{3.1}$$

$$\hat{y} \sim condition + time + (1|participant) \tag{3.2}$$

$$\hat{y} \sim condition * time + (1|participant) \tag{3.3}$$

To test whether the models improved by adding predictors we used likelihood ratio testing. If the model fit did not improve by adding the predictor 'time', results from the first model were reported.
A linear mixed model (LMM) was estimated for the ERP and TRF amplitudes, for each subjective workload question, and for the time to complete the surgery task. As the number

of mistakes during the surgery task (zero to three mistakes) and the number of times the tissue was damaged provided count data, a generalized linear model (GLM) with a poisson distribution was estimated. The LMM and GLM were estimated using the R package lmer4 (v. 1.1-30). Fixed effects were evaluated using Satterthwaite approximations within the R package lmerTest, which estimates the degrees of freedom to calculate p-values.
For the memory task, memory scores were first calculated using edit distance scoring (Gonthier, 2022) (i.e., scores range from 0 to 1). Then, the scores received ordinal values which were fit using a cumulative link mixed model using the clmm function of the ordinal package (v. 12-4).
Evidence for an effect was assumed for α =.05. We corrected for multiple comparisons using α =.05/3 = 0.017 for the ERP/TRF amplitudes (i.e., N1, P2, N2), the comparison between prediction values of the acoustic features (i.e., envelope, noise-reduced envelope, onsets), subjective workload questions (i.e., effort, frustration and distraction), and surgery task performance (i.e., duration, mistakes, and tissue damage).

3.4 Results

For each response we evaluated a condition difference using regression models. We explored whether adding time as a predictor improved the model fit. Table 3.1 shows the selected model for each response and whether the corresponding beta values were significant.

3.4.1 Subjective measures of demand

After each block, we asked participants how effortful, frustrating, and distracting they perceived the task. There was a significant increase in scores from the low- to high-demand condition for all three measures (Fig. 3.2a-c; effort: $b = 3.73$, $SE = 0.22$, $p < .001$; frustration: $b = 2.71$, $SE = 0.28$, $p < .001$; distraction: $b = 1.31$, $SE = 0.19$, $p < .001$). There was no significant change in theses measures over time, i.e., for all three questions the model fit did not improve by adding time as a predictor (Supplementary Table S3.2).

3.4.2 Objective measures of performance

The memory score, which could range between 0 to 1, was in the low-demand condition on average 0.97 (SD=0.13) and in the high-demand condition 0.75 (SD=0.24). The first model, which included only condition as a predictor, was used (Supplementary Table S3.3) and revealed a significant decrease in memory score from the low- to high-demand condition (Fig. 3.2d, $b = -4.63$, $SE = 0.39$, $p < .001$).
To evaluate the performance during the surgery task we used three measures: task duration (i.e., time to finish the surgery task), the number of mistakes (i.e., dropping or rupturing

a vessel), and the number of times the tissue was damaged (i.e., tissue that was touched with either of the tools). Figure 3.2e-g illustrates the results of the behavioral performance. The performance was already high at the beginning of the experiment, which is evident in the low number of mistakes in most of the blocks. The average number of mistakes is 0.33 (SD=0.48) in the first block and 0.18 (SD=0.39) in the last block. The number of times tissue was damaged changed from 6.2 (SD=3.96) in the first block to 5.4 (SD=4.88) in the last block. Furthermore, task duration appeared to improve, particularly in the first ten blocks, after which it remained relatively stable. For neither of these measures, we found a significant difference between conditions (task duration: $b = -1.842$, $SE = 1.57$, $p = .228$; mistakes: $b = 0.14, SE = 0.19, p = .445$; tissue damage: $b = -0.02, SE = 0.034, p = .556$). However, for all measures adding time as a predictor significantly improved the model fit (Supplementary Table S3.3) and revealed a significant decrease over time (task duration: $b = -0.91, SE = 0.1, p < .001$; mistakes: $b = -0.03, SE = 0.01, p = .0046$; tissue damage: $b = -0.01, SE = 0.02, p < .001$), in other words, participants became faster and made fewer mistakes.

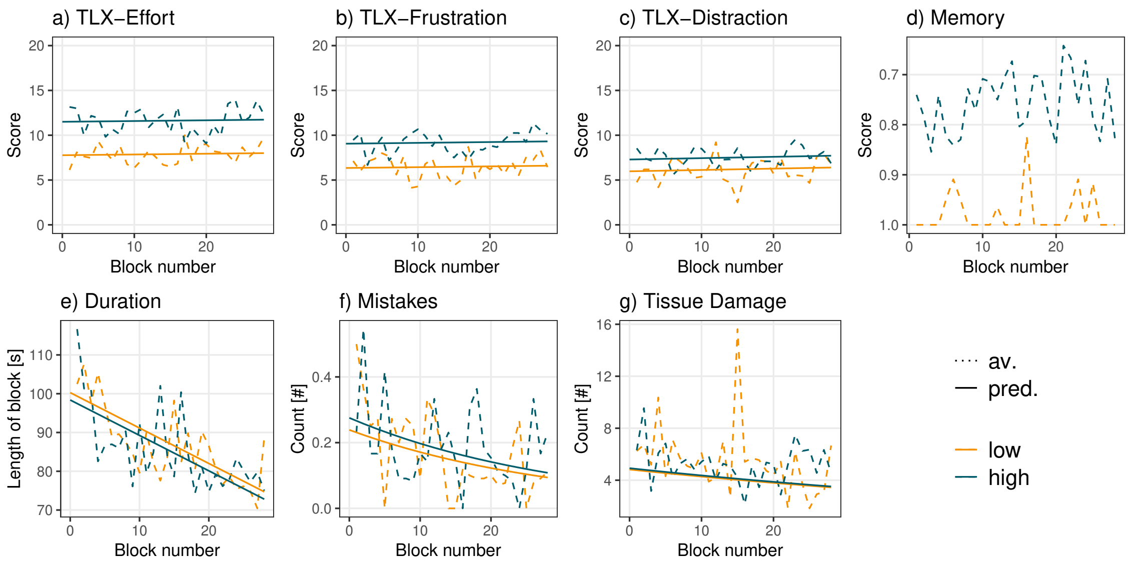


Figure 3.2: (a-c) Subjective responses. (d) Memory task performance. (e-f) Surgery task performance. Dotted lines show averaged data over participants for each condition and block. Solid lines represent predicted responses that were calculated using the second model. Predicted responses for the memory task were not computed as a cummulative link model was used.

3.4.3 Demand and time-on-task effects for ERPs in response to the spoken letters.

The response to the task-irrelevant spoken letters was evaluated by averaging GED component amplitudes for the N1 (Fig. 3.4a), P2 (Fig. 3.4b), and N2 (Supplementary Figure S3.4a) over their respective time-window. Both the morphology as well as the topographies of the GED components with the strongest eigenvalues (Supplementary Figure S3.4a) are physiologically plausible.

Contrary to our expectation, we observed no significant effect of demand on the N1 ($b = 0.4, SE = 1.84, p = .827$), the P2 ($b = 0.21, SE = 1.91, p = .256$), or the N2 ($b = 0.169; SE = 2.05; p = .411$) amplitude. Exploring time as a predictor significantly improved the models for the N1 and P2 amplitudes but not for the N2 amplitude (Supplementary Table S3.3) and revealed a significant N1 amplitude decrease ($b = 0.3, SE = 0.11, p = .0012$) and a P2 amplitude increase ($b = 0.38, SE = 0.11, p = .0014$) over time.

3.4.4 Prediction accuracies for the TRFs in response to the OR playback.

The TRFs were calculated by relating the continuous OR playback to the EEG signal. We used three acoustic features to calculate the TRF, namely the envelope extracted from the raw audio, the envelope extracted from the noise-reduced audio, and the onsets extracted from the raw audio (Fig. 3.3a). To compare these features, we calculated the prediction accuracy for unseen neural data. Overall, the prediction values were small (Fig. 3.3b) which is common for this measure. Importantly, the envelope of the noise-reduced audio as well as the onsets showed significantly higher prediction accuracies compared to the envelope from the raw audio (noise-reduced envelope: $W = 3, p < .001$; onsets: $W = 1, p < .001$). The noise-reduced envelope and onsets did not significantly differ from each other ($W = 118, p = .799$).

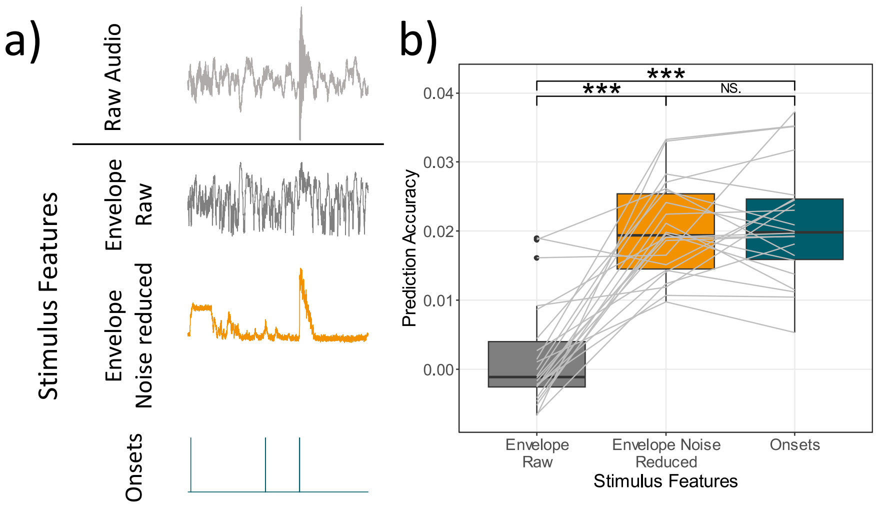


Figure 3.3: a) The acoustic features used for TRF model estimation. The same stimulus snippet is shown as the raw audio, envelope of the raw audio, envelope of the noise-reduced audio, and onsets of the raw audio. b) Prediction values of each acoustic feature. Each line represents the change in prediction value between acoustic features for each participant. *** $p < .001$

3.4.5 Demand and time-on-task effects for TRFs computed from the noise-reduced envelope and onsets.

We computed the TRF in response to the noise-reduced envelope (hereafter $\text{TRF}_{\text{envelope}}$) and in response to the onsets (hereafter $\text{TRF}_{\text{onsets}}$). Although there was no difference between prediction accuracies between these two acoustic features, they might still represent different aspects of the soundscape and, therefore, different responses. To enhance the signal-to-noise ratio and to spatially filter the data GED components were extracted. The mean amplitudes over the N1, P2, and N2 time-windows were used for evaluation. Figure 3.4c and d shows the distinct time-series and topographies for the N1 $\text{TRF}_{\text{envelope}}$ and the N1 $\text{TRF}_{\text{onsets}}$, respectively. The remaining GED time-series, as well as GED eigenvalues can be found in supplementary figure S3.3 and S3.4. Visually there are considerable differences in the temporal evolution between the $\text{TRF}_{\text{envelope}}$ and the $\text{TRF}_{\text{onsets}}$. While the $\text{TRF}_{\text{envelope}}$ follows a trajectory similar to the ERPs its response peaks occur earlier than that of the ERPs and the $\text{TRF}_{\text{onsets}}$.

Regarding the N2 $\text{TRF}_{\text{envelope}}$, the GED component with the second strongest eigenvalue showed more plausible trajectories and topographies than the first GED component (Supplementary Figure S3.4).

As $\text{TRF}_{\text{envelope}}$ and $\text{TRF}_{\text{onset}}$ showed similar prediction values, we evaluated amplitudes of both TRFs, and corrected for additional multiple comparisons using α =.05/3 (responses) / 2 (acoustic feature) = 0.0083. For all $\text{TRF}_{\text{envelope}}$ amplitudes, we found no condition difference (N1: $b = -0.94, SE = 0.57, p = .102$; P2: $b = -0.37, SE = 0.49, p = .449$; N2: $b = 0.75, SE = 0.45, p = .095$) and adding time as a predictor did not lead to a better model fit compared to the first model (Supplementary Table S3.3).

Regarding the $\text{TRF}_{\text{onsets}}$ amplitudes we found no condition differences (N1: $b = 0.49, SE = 0.41, p = .23$; P2: $b = 0.025, SE = 0.39, p = .948$; P2: $b = -0.12, SE = 0.35, p = .716$). Adding time as a predictor yielded a better model fit for the N1 amplitude, but not for the other amplitudes. This revealed a trend towards significance for the predictor time for the N1 $\text{TRF}_{\text{onsets}}$ ($b = 0.07, SE = 0.02, p = .011$).

3.5 Discussion

In this study, we examined the impact of varying demand and time-on-task on subjective responses, task performance, and EEG responses within a naturalistic setting, utilizing a laparoscopic simulator and an operating room soundscape. We found divergent subjective and objective behavioral responses. Participants perceived the high-demand condition as more demanding than the low-demand condition, which remained stable over time. However, surgery task performance did not differ between demand conditions but demonstrated improvement over time.

The overall perceived demand of the two conditions was evaluated with subjective work-

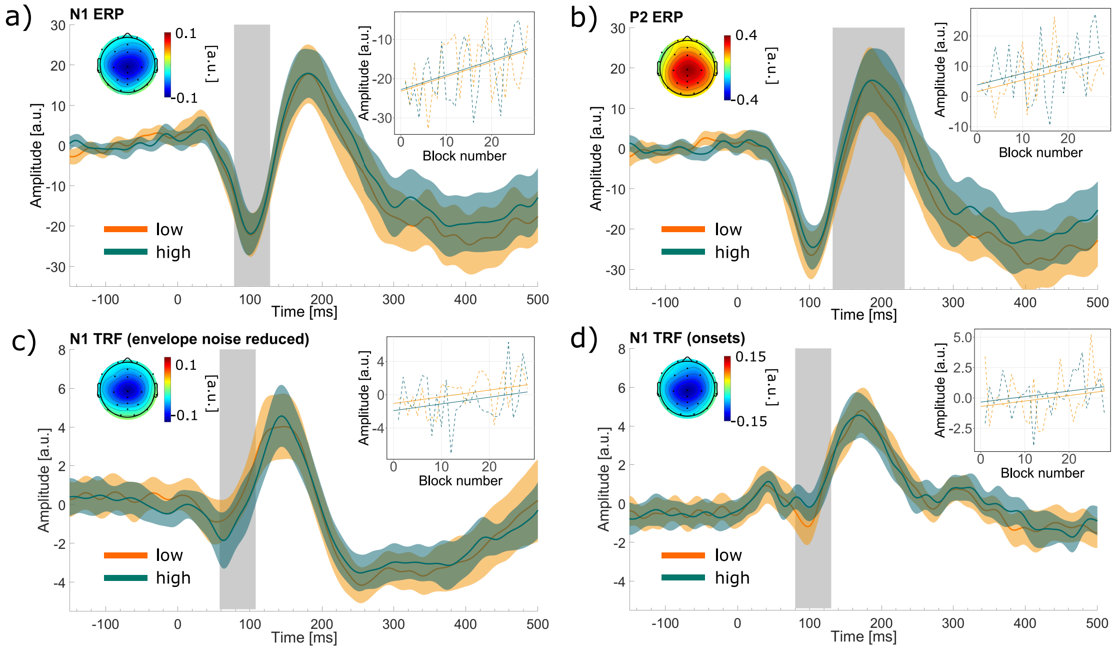


Figure 3.4: The first GED components of the (a) N1 ERP (demand: n.s., time: $p = .0012$), (b) P2 ERP (demand: n.s., time: $p = .0014$), (c) N1 $\text{TRF}_{\text{envelope}}$ (demand: n.s., time: n.s.), and (d) N1 $\text{TRF}_{\text{onsets}}$ (demand: n.s., time: n.s.). The morphology shows the GED time-series for each condition, averaged over participants. The shaded area represents the confidence interval across participants. The grey bar shows the time-window that was used to derive the amplitudes. The topographies depict the forward model of the GED weights and shows which channels contributed most to the GED component. The graph to the right, shows the change in amplitude across time and conditions. Dotted lines show averaged data over participants for each condition and block. Solid lines represent predicted responses that were calculated using the second model.

load questions. The observed effect that high demand compared to low demand increases subjective workload ratings during a surgery task aligns with prior research (Gao et al., 2019; Modi, Singh, Darzi, & Leff, 2020; Zander et al., 2017). A possible explanation for the perceived differences in demand across conditions, despite consistent performance levels across conditions, is that participants may have prioritized their performance of the surgery task. By putting more cognitive effort into the surgery task in the high-demand condition, they were able to achieve similar surgery task performance across both conditions. Likewise, a surgeon would prioritize performance during an actual surgery for the benefit of the patient, even if it means an increase in perceived demand. This implies that task performance and subjective experience are not necessarily related and may also explain the contradictory findings reported in the literature. Some studies report a negative impact of increased demand on performance (Gao et al., 2019; Moorthy, Munz, Dosis, et al., 2003), while others show no effect (Hsu, Man, Gizicki, Feldman, & Fried, 2008), or suggest that the effect depends on the investigated performance measure (Modi et al., 2020; Poolton et al., 2016). Most studies, did not assess the subjective demand, and only concentrate on task performance. However, this limits our understanding how perceived demand and surgical task performance relate and should be addressed in future research by incorporating subjective demand measures.

Regarding the impact of time-on-task on performance, previous research has also demon-

	Response	Model	Condition	Time
subjective workload questions	tlx-effort	1	3.73*	
	tlx-frustration	1	2.71*	
	tlx-distraction	1	1.31*	
behavioral performance	memory score	1	-4.63*	
	duration	2	-1.842	-0.91*
	mistakes	2	0.14	-0.03*
	tissue damage	2	-0.02	-0.01*
ERPs	N1	2	0.4	0.3*
	P2	2	0.21	0.38*
	N2	1	0.169	
$TRF_{envelope}$	N1	1	-0.94	
	P2	1	-0.37	
	N2	1	0.75	
TRF_{onsets}	N1	2	0.49	0.07 .
	P2	1	0.025	
	N2	1	-0.12	

Table 3.1: The table shows the selected model for each subjective, behavioral, and neural response. ’Model’ lists the chosen model for each response. Model 1 included ’condition’ as a predictor and model 2 ’condition’ and ’time’ as predictors. Beta values are listed below each predictor. The stars indicate predictors below the Bonferroni-corrected p-value. A trend is marked with a dot.

strated that task performance improves over time (Hu, Lu, Tan, & Lomanto, 2016; Maimon et al., 2022; Suárez et al., 2022; Zakeri, Mansfield, Sunderland, & Omurtag, 2020), highlighting visible training effects within a short amount of practice when using the same surgery task. We did not find a change in perceived demand over time, similar to Suárez et al. (2022) who found that training did not reduce demand for the same NASA-TLX items as used in this study (i.e., effort and frustration) or the total score. Others reported that surgical training lowers the NASA-TLX total score (Hu et al., 2016; Zakeri et al., 2020). This once again demonstrates that the relationship between task performance and perceived demand is not straightforward. The use of various methods to vary demand and surgical tasks, each with different performance parameters and subjective and objective measures, also complicates the generalization of findings across studies (Georgiou, Larentzakis, & Papavassiliou, 2017). To improve generalization, it would be useful to evaluate how the demand experienced during actual surgeries reflects the demand experienced during surgical simulations. Our results suggest that situational demand, rather than actual surgical skill, influences the perceived demand in inexperienced participants. This suggests that surgical training programs should integrate high-demand simulations to more effectively prepare novice surgeons, who are particularly susceptible to changes in task demand (Arora et al., 2010; Hsu et al., 2008), for the cognitive and emotional demand they will encounter in real surgeries. By systematically exposing novice surgeons to challenging scenarios, training programs can enhance their ability to manage stress and maintain performance under pressure, potentially reducing errors and improving patient outcomes. Furthermore, novice surgeons may benefit from training to reflect on their individual susceptibility to noise.

This is because noise reduction interventions are especially beneficial for surgeons who perceive the soundscape as disturbing (Engelmann et al., 2014). Such training could also incorporate communication strategies to reduce avoidable noise sources in the operating room, especially at the start of critical and high-demand phases of the surgery. This would help to reduce the overall experienced demand for the surgeons.

As an alternative objective measure to capture the perceived task demand, we investigated two neural response measures to the task-irrelevant soundscape. Specifically, we analyzed ERPs in response to spoken letters and TRFs in response to the OR playback. In general, we find a clear neural response to the letters as well as for the naturalistic OR playback, which is in line with our previous study showing significant responses for naturalistic soundscapes (Rosenkranz et al., 2023). This further expands the use of mobile EEG in applied settings (Wascher et al., 2021), as both measures can be used to study the neural response to a complex natural soundscape in a work-like environment. Contrary to our hypothesis, we found no significant effect of task demand on the neural measure of sound processing, neither in response to the letters nor to the OR playback. Participants' subjective ratings indicate that our intended manipulation — to increase the scenario's challenge — had an effect, with participants perceiving the soundscape as more distracting in the high-demand condition. However, this perceptual shift is not reflected in our neural measurements. Our exploratory analyses revealed time-on-task effects, for the N1 and P2, and a trend for the TRF_{onsets} N1. This was apparently unrelated to the subjectively perceived distraction, which remained stable over time.

Our results should be considered in light of the neural measures used to investigate different aspects of the soundscape, i.e., the ERPs and TRFs. The ERPs captured responses to the regularly occurring letters associated with the memory task. To increase the distracting nature of the task-irrelevant soundscape and simulate demanding internal cognitive processes we adapted a serial-recall paradigm. We observed no effect on any ERP time-window, however, previous studies using a serial-recall paradigm have shown that an increase in distraction leads to an increase in N1 ERP amplitude (Campbell, Winkler, Kujala, & Näätänen, 2003). In contrast to a classic serial-recall paradigm, we used the serial-recall task as a secondary task with a long retention interval (more than 60 seconds), and presented distractors at a rate of 3 seconds. Furthermore, we presented the letters in the presence of background noise, which may have reduced distraction effects in our paradigm (Haapakangas et al., 2020). Our choice of distractor presentation (long inter-trial-intervals and in the presence of noise), in combination with an engaging surgery task, may have enabled participants to easily ignore the letters over time, independent of the number of items to recall, resulting in similar responses across conditions. Notably, the results showed a decrease in N1 amplitude and an increase in P2 amplitude, indicating neural adaptation due to repeated stimulus presentation. This may be similar to the adaptation observed for brief repetitive tones (Hari, Sams, & Järvilehto, 1979; Wong, Huo, & Maurer, 2023). The decrease in ERP N1 amplitude could indicate that distraction due to the letters decreased (Campbell et al., 2003), independent of the conditions. Participants may have initially attended to the letters, e.g., due to their deviation from the continuous soundscape, and learned that the letters are task-irrelevant and can be ignored as such (Schröger, 1997). The P2 amplitude increase is more difficult to interpret, as only few studies investigated it in the context of distraction. One study showed an amplitude increase with increased distraction (Regenbogen et al., 2012), which would contradict our interpretation, while another study found no effect of distraction on P2 amplitude (Mahajan, Kim, & Davis, 2020). In summary, the results suggest that the task-irrelevant spoken letters did not sufficiently distract participants to elicit distinct neural responses under varying demands. However, the changes in amplitude across time indicate that the responses can capture some form of modulation. As this was an exploratory finding, further investigation is necessary to reveal how responses to task-irrelevant acoustic events change over time when presented in a realistic work-like setting. This information could shed light on how the brain adapts to such events in real life.

The TRFs captured responses to the realistic OR playback. To the best of our knowledge this is the first study investigating demand effects on responses to task-irrelevant ongoing soundscapes. Despite the fact that we measured a clear TRF to the soundscape, we did not find a difference between conditions. One explanation for this null-finding is that TRFs calculated from the entire soundscape are not sensitive enough to detect demand effects. A small portion of sounds may have elicited demand effects similar to ERPs (Brockhoff et al., 2022). However, these effects may have been reduced by mostly demand-unrelated responses. Support for this explanation comes from our previous study (Rosenkranz et al., 2023), where we found attention-dependent responses for specific highly salient sounds, but only marginal differences in the TRF to the background soundscape. This might also explain why we found no time-on-task effects for the TRF, except for the trend of a decrease for the TRF N1 calculated from the onsets (TRF_{onsets}), despite the effect for the ERPs. While the ERPs captured responses to similar sounds, the TRFs captured responses to the OR playback which was much more diverse than the letters. This might have decreased a potential time-on-task effect for the TRFs compared to ERPs, as different and overlapping responses are less sensitive to detect effects compared to similar and isolated responses (Haupt et al., 2024). For example, an audio signal, that contains a constant high level of noise, would result in an essentially flat envelope, for which no reliable TRF could be computed. Any acoustic event, that is embedded in this background sound, though perceivable to the human ear, would not be reflected in the envelope. This is evident in our finding that TRFs computed from the raw envelope, which captured little variation of the soundscape, provided low prediction values. The individual acoustic events, such as the beeps of machinery or the clatter of tools, were only adequately captured by reducing the noise in the raw envelope. By highlighting the individual acoustic events, the envelope became more similar to the computation of the onsets. The similarity between these two features is expressed in similar prediction values. This suggests that the brain's processing of sounds goes beyond simply following the acoustic envelope (Di Liberto, O'Sullivan, & Lalor, 2015; Drennan & Lalor, 2019). Instead, it actively distinguishes and monitors individual acoustic events in the presence of background noise, similar to how speech is perceived in noisy environments (Khalighinejad, Herrero, Mehta, & Mesgarani, 2019).

Note, that similar prediction values, do not necessarily imply a similar neural representation of the soundscape (Haupt et al., 2024). This is hinted by the trend in the TRF_{onsets} N1 which was not observed in the $TRF_{envelope}$. However, the similar prediction accuracies indicate that onset information about acoustic events is enough to estimate the neural response to complex soundscapes. This can have practical implications when privacy is important (e.g., in an actual OR context where medical information about a person are discussed). It would be sufficient to extract acoustic onsets automatically, without the need to record the raw audio (Hölle & Bleichner, 2023b; Hölle et al., 2022). Overall, we have demonstrated that it is possible to obtain responses from continuous task-irrelevant soundscapes. However, future studies may benefit from describing the soundscape not only in the form of an envelope or by extracting onsets, but to differentiate between different sound sources to obtain a more fine grained picture.

An alternative explanation for the discrepancy between our objective measures (i.e., task performance and neural measures) and the subjective reports is that the subjective reports did not accurately reflect participants' perceived demand throughout an entire block. As the subjective responses were collected at the end of a block, they may reflect the perceived demand of the memory task during encoding. In other words, participants found it difficult to remember two compared to eight letters, but experienced the surgery task as similar demanding across conditions. It is important to note that although performance on the memory task differed between conditions, the difference was generally small. This may be attributed to the homogeneity of the participant pool and their high cognitive capacities, comprising medical students of a similar young age. Although participants still rated the high-demand condition to be more demanding, the high-demand condition has not depleted their cognitive resources enough, which could explain the absence of differences in our objective measures. Furthermore, given the complexity of the task, accurately pinpointing the source and amount of demand — whether it's remembering the letters, the surgery task, or a combination of both — may have been difficult for participants. Indeed, the NASA-TLX has been criticized that different participants might rate different aspects of the task (McKendricka & Cherry, 2018). Collecting separate subjective responses for each task may have improved interpretation. However, in natural environments, the source of demand, whether it is the environment or the task, can rapidly change, making subjective responses at a single point in time difficult to interpret. This once again emphasizes the importance of continuously and objectively monitoring sources of demands, such as the soundscape, to understand how they affect the individual.

In order to improve the understanding of varying demands on responses to ongoing, task-irrelevant soundscapes, we offer three suggestions for future work. Firstly, a different manipulation of demand could clarify whether neural responses to irrelevant sounds are indeed insensitive to detect demand effects in complex settings. The measurement of EEG during a standing, bi-manual task presents a greater challenge than in traditional stationary setups. Such complex scenarios also engage a variety of cognitive processes. These two aspects likely increase individual variability in EEG responses and reduce sensitivity to isolated demand effects. Furthermore, it is possible that the demand manipulation may have not been strong enough to elicit differences in the neural response. For these reasons, it might be necessary to implement manipulations with larger differences in cognitive demand. Secondly, we used a general representation of the soundscape, by computing the envelope and onsets of the playback. A more detailed description of the soundscape, for example, by differentiating between specific sounds may lead to a better estimate of the neural response (Haupt et al., 2024). Finally, our study involved a small sample of young medical students with limited OR experience. This offers valuable insights into the feasibility of using EEG in complex settings. However, the generalizability to the clinical personal is limited. Future studies would benefit from increasing heterogeneity and sample size. Tailoring the sample to the target population by including experienced surgeons would provide deeper understanding of how personnel accustomed to complex environments process and perceive the soundscape under varying demand. Implementing these suggestions will advance the application of EEG in work environments and enhance our understanding of whether and how cognitive states can be inferred from responses to natural, task-irrelevant soundscapes.

Our ultimate goal is to study neural responses in real-world settings and workplaces, such as the OR. These settings are characterized by a complex and varying soundscape, while individuals engage in tasks of varying complexity, which the researchers have limited control over. In order to learn from our results for other studies that are interested in real-world recordings, it is important to discuss some properties of the soundscape we have used here. We used the playback of a natural soundscape that was recorded in an OR using a static microphone positioned at the center of the room. We presented this soundscape via two loudspeakers to a stationary participant. This setup ensured that the soundscape maintained realistic spatial properties from the listener's perspective, reflecting the acoustic environment of an operating room. Despite the naturalness of this soundscape, there are important differences between the OR playback we used and an actual OR soundscape experienced in real-life situations. One key difference is sound expectancy. For instance, in an OR setting, there is an expectation that specific actions one can see, such as laying down a tool, will produce a particular sound. In other words, there is a congruence between visual and auditory information. In our study, this audio-visual congruence is absent; all the sounds our participants hear are not congruent with the actual situation they are in. One might expect these differences to affect the neural response we have measured, although the direction of this effect is difficult to predict. Some speech tracking studies have shown an enhanced tracking of the speech envelope when congruent visual information is provided (Crosse, Butler, & Lalor, 2015). On the other hand, a mismatch of visual and auditory information could also lead to a larger neural response (Ullsperger, Erdmann, Freude, & Dehoff, 2006). A second difference between the soundscape experienced in the OR and our OR playback is the relevance of the sounds for the participant. In an OR setting, many sounds carry contextual information, for example, the sound of an alarm, the feedback sound of a tool that is in use, or relevant communication between personnel, that are critical to the surgery and sometimes require task-related actions. In contrast, the soundscape we presented was entirely task-irrelevant for our participants, essentially transforming all sounds into a meaningless noise background. One can expect that for personally relevant or salient information within a soundscape influences its processing and thereby the neural response (Holtze et al., 2021; Roye, Jacobsen, & Schröger, 2013; Straetmans et al., 2021). Taken together, these factors may significantly change the neural response to the naturalistic soundscape we have presented, compared to when this naturalistic soundscape would have been experienced in its actual context. Moving forward, future studies should be aware of these important factors when studying naturalistic soundscapes. To conclude, our results demonstrated the feasibility and potential of combining mobile EEG and audio recordings to study cognitive processes in complex, real-world environments. Our findings showed a temporal change in neural responses to specific sounds, but not to the entire soundscape. This suggests that responses to isolated sounds may be more sensitive indicators of alterations in auditory processing than responses to the entire soundscape. Furthermore, we observed that increased task demand resulted in a higher perceived demand and distraction, but these changes were not reflected in the neural responses or objective task performance measures. This discrepancy indicates that a high performance can be maintained even in demanding environments, but potentially at a cost to a person's wellbeing. These insights have important implication for the surgical field and other workplaces. They highlight the need to consider both subjective experiences and objective measures when evaluating cognitive demand in high-stakes environments. Our approach contributed to methodological advancements in neurophysiological research and opens new avenues for optimizing auditory environments in surgical workplaces, ultimately contributing to both performance and the wellbeing of personnel.

3.6 Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

3.7 Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) ID 411333557 & 490839860, and by the Forschungspool Funding of the Oldenburg School of Medicine and Health Science. We thank Agnes Hanssen for her help in data acquisition.

3.8 Additional information

The author(s) declare no competing interests. During the preparation of this work the authors used ChatGPT (v. 3.5) and DeepL Write (acadamic style) to improve grammar and wording. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

3.9 Supplementary material

Deviations from preregistration

TRF analyses

- Initially we ran the cross-validation procedure 100 times and randomly chose 90% of the blocks as test data and the remaining 10% of the blocks as training data. However, we realized that this is computationally expensive while resulting in very similar prediction values as with the approach in the manuscript.
- During the last step of TRF calculation we computed the window from -220 to 500 ms. In the preregistration the window ranged from -200 to 500 ms, however, due to an early artefact common to TRF calculations (Crosse et al., 2021) we had to increase the window.
- Due to an oversight we did not detail in the preregistration how we determined the individual optimal lambda value.
- We z-standardized the TRF values before GED analyses. This was necessary as the individual lambda range was large, thus resulting in individual TRFs on different scales which lead to violations of the normality assumption for the statistical comparison.

GED analyses Our initial idea was to individualize the analyzes pipeline, especially the GED computation and compute a spatial filter per participant. However, it turned out that this worked well for our pilot data but to a much lesser degree for the entire dataset. The GED analyses led to component time-series and maps that were hard to interpret or not meaningful making component selection difficult. Therefore, we used a grand average GED model and made the following adaptations:

- We used the GED weights calculated from all subjects and applied them to each subject and block. The initial idea was to calculate GED weights for each participant calculated from the grand average over blocks and apply the weights to all blocks of of this participant.
- We calculated fixed time-windows to determine the component time-windows. Initially, the full width at half maxmimum/minimum with respect to the peak was used, but this lead to time-windows that extended over the response of interest.
- We calculated the S and R covariance matrix for each participant and cleaned the matrix before averaging. This is a step included in Cohen (2022)(Cohen, 2022). Initially, we used the grand average ERP of a participant to compute one S and one R matrix, thereby making the cleaning step unnecessary.

Further adaptions to the GED analyses were made:

- Initially the peak of all ERP and TRF components were searched for in the same time-window. However, it turned out that the envelope response peaks earlier than the ERPs and TRFs calculated from the onsets due to temporal smearing. To account for this, we searched for the envelope peaks in an earlier time-window, as stated in the manuscript.
- In the preregistration the lambda value to regularize the R matrix was defined as $\lambda = 0.1$ which was a typo. The initial analyses was set up with $\lambda = 0.01$ as stated in the manuscript and which was also used in Cohen (2022) (Cohen, 2022).

Statistical analyse

- The initial idea was to calculate the maximum LMM (i.e., $\hat{y} \sim condition + (condition|participant)$). However, in most cases this model did not converge, especially when adding time as a predictor. We decided to start with model 1, then successively add fixed effects, and finally evaluate the change in model fit.
- For the memory scores, we initially wanted to calculate a beta regression model, but realized that these are not applicable when responses are bound to 0 or 1, which was the case for our data.

Shift of letter onset

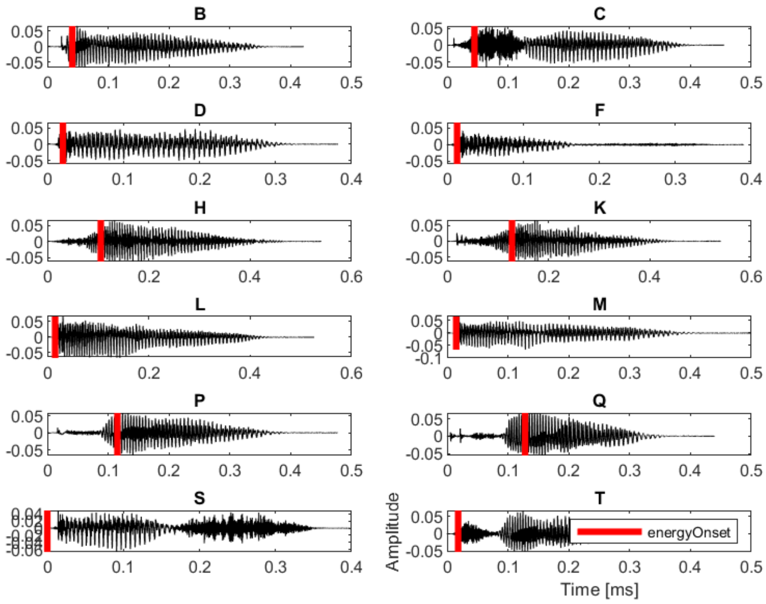


Figure S 3.1: Time-representation of each letter. The red line indiactes the shift of the letter onset.

Removed IC containing muscle movement

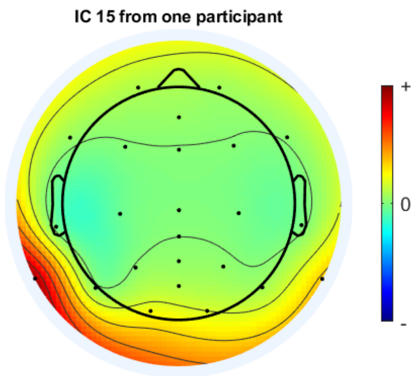
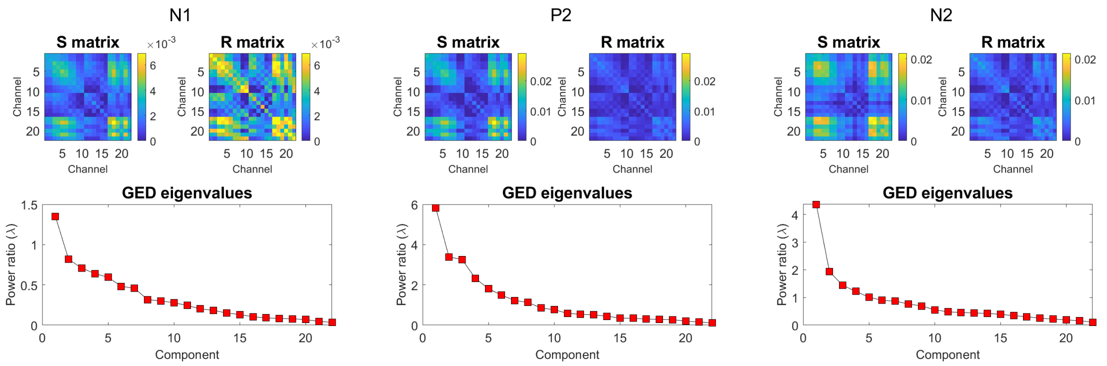
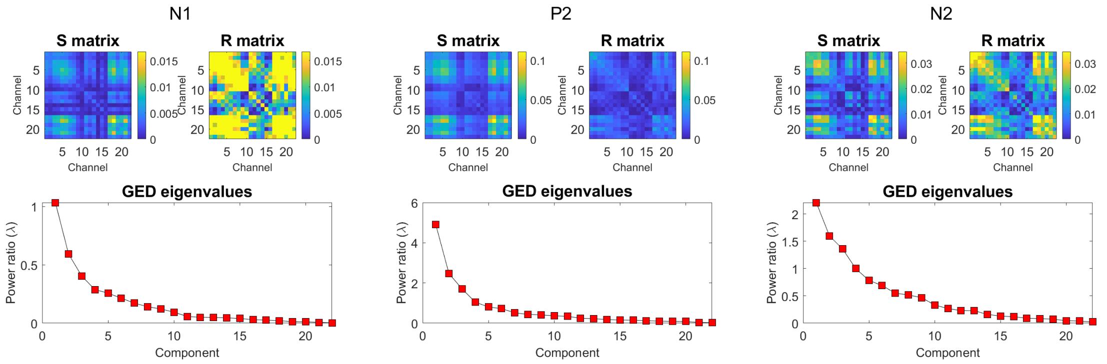


Figure S 3.2: This is an example of an independent component of one participant that was manually removed, as it contained noise at the mastoid electrode.

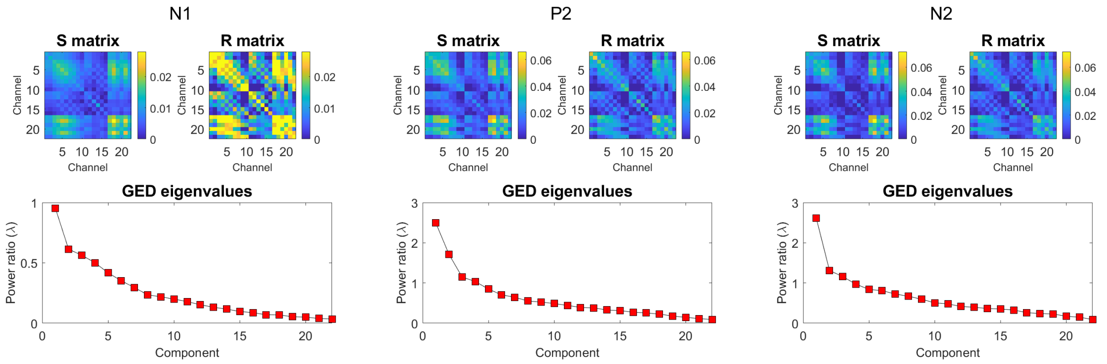
GED covariance matrices



(a) ERP components in response to the letters.



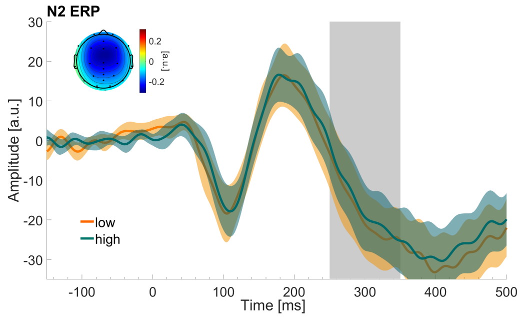
(b) TRF components calculated from the noise reduced envelope.



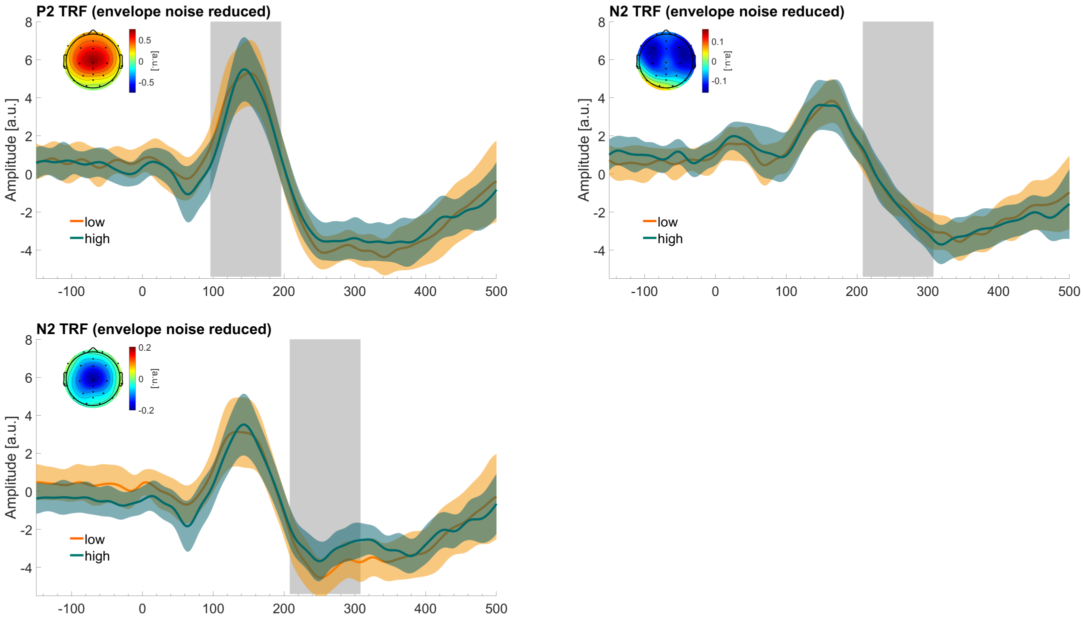
(c) TRF components calculated from the osets.

Figure S 3.3: GED covariance matrices and eigenvalues for (a) ERPs, and TRFs calculated from the (b) noise reduced envelope and (c) onsets. Upper left and right plot depict the $\bar{S}$ and $\bar{R}$ covariance matrix, respectively. The lower plot depicts the eigenvalues of each component.

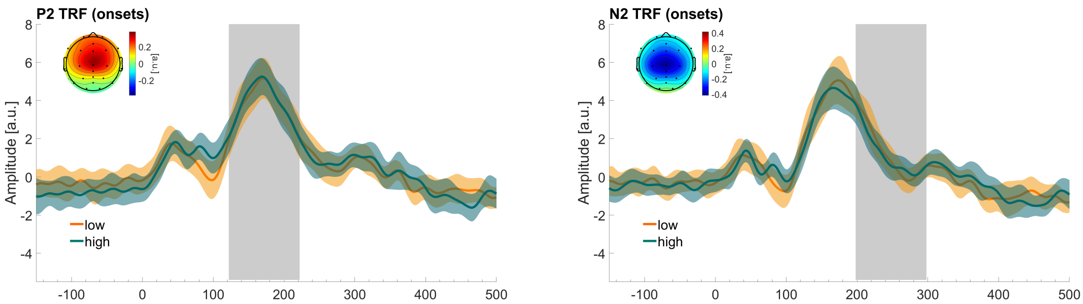
GED components



(a) ERP N2: First GED component.



(b) **(left top)** First GED component of P2 TRF envelope (noise-reduced). **(right top)** First GED component of N2 TRF envelope (noise-reduced) and **(left bottom)** second GED component of N2 TRF envelope (noise-reduced). The second component was chose for the N2 analyses, as the map and topographies look more plausible.



(c) **(left)** First GED component of P2 TRF onset. **(middle)** First GED component of N2 TRF onset.

Figure S 3.4: This figure shows GED components that were analyzed but not shown in the manuscript. The GED time-course displays averaged data (line) and standard error (shaded area) across participants. The grey area shows the time-window that was used to compute the average amplitude. The topographies depict the forward model of the GED component.

Statistical model comparisons

Behavioral and subjective Results

response	predictor	AIC	BIC	logLik	Chisq	Df	Pr(>Chisq)
tlx-effort	condition	3164.4	3182.1	-1578.2			
	condition+time	3166.0	3188.2	-1578.0	0.3593	1	0.5489
	condition*time	3166.7	3193.3	-1577.3	1.3156	1	0.2514
tlx-effort	condition	3440.7	3458.5	-1716.4			
	condition+time	3442.5	3464.7	-1716.2	0.2661	1	0.6059
	condition*time	3444.4	3471.0	-1716.2	0.0961	1	0.7565
tlx-distraction	condition	3006.0	3023.7	-1499.0			
	condition+time	3006.4	3028.7	-1498.2	1.5193	1	0.2177
	condition*time	3007.8	3034.5	-1497.9	0.6005	1	0.4384
Performance time	condition	5656.2	5674.0	-2824.1			
	condition+time	5576.4	5598.6	-2783.2	81.8695	1	<2e-16 ***
	condition*time	5577.7	5604.3	-2782.8	0.6821	1	0.4089
Mistakes	condition	619.97	633.29	-306.98			
	condition+time	613.99	631.74	-302.99	7.9818	1	0.004725 **
	condition*time	614.66	636.86	-302.33	1.3249	1	0.249721
Tissue damage	condition	3807.9	3821.2	-1901.0			
	condition+time	3778.8	3796.6	-1885.4	31.0797	1	2.476e-08 ***
	condition*time	3778.8	3801.0	-1884.4	1.9995	1	0.1573

Table S 3.2: The table shows model performance of the statistical models. The simplest model contained condition as a predictor. There is no model comparison for the memory condition, as the models did not converge when adding time as a predictor. Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

EEG data

response	predictor	AIC	BIC	logLik	Chisq	Df	Pr(>Chisq)
ERP - N1	condition	5484.0	5501.5	-2738.0			
	condition+time	5475.5	5497.4	-2732.7	10.4946	1	0.001197 **
	condition*time	5476.2	5502.6	-2732.1	1.2575	1	0.262118
ERP - P2	condition	5549.9 5567.5	-2770.9	5541.9			
	condition+time	5541.7	5563.7	-2765.9	10.1876	1	0.001414 **
	condition*time	5543.3	5569.6	-2765.6	0.4539	1	0.500484
ERP - N2	condition	5622.4	5639.9	-2807.2			
	condition+time	5623.6	5645.6	-2806.8	0.7320	1	0.3922
	condition*time	5625.5	5651.9	-2806.8	0.0959	1	0.7568
TRFenv - N1	condition	4067.9	4085.5	-2030			
	condition+time	4066.6	4088.6	-2028.3	3.2459	1	0.0716 .
	condition*time	4068.2	4094.6	-2028.1	0.4148	1	0.5196
TRFenv - P2	condition	3887.3	3904.9	-1939.4			
	condition+time	3889.2	3911.2	-1939.6	0.0451	1	0.8318
	condition*time	3891.2	3917.6	-1938.6	0.0131	1	0.9090
TRFenv - N2	condition	3778.3	3795.9	-1885.2			
	condition+time	3780.3	3802.3	-1885.1	0.0175	1	0.8946
	condition*time	3782.3	3808.6	-1885.1	0.0258	1	0.8723
TRFons - N1	condition	3663.6	3681.2	-1827.8			
	condition+time	3659.2	3681.2	-1824.6	6.3816	1	0.01153 *
	condition*time	3661.2	3681.5	-1824.6	0.0408	1	0.83994
TRFons - P2	condition	3610.4	3628.0	-1801.2			
	condition+time	3610.1	3632.1	-1800.1	2.2199	1	0.1362
	condition*time	3610.6	3637.0	1799.3	1.5038	1	0.2201
TRFons - N2	condition	3440.1	3457.7	-1716.0			
	condition+time	3439.8	3461.8	-1714.9	2.2371	1	0.1347
	condition*time	3441.8	3468.2	-1714.9	0.0232	1	0.8790

Table S 3.3: The table shows model performance of the statistical models. The simplest model contained condition as a predictor. TRFenv: TRFs were calculated using the noise reduced envelope. TRFons: TRFs were calculated using the onsets of the raw audio. Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

4 Study III - The effect of task demand on EEG responses to irrelevant sound and speech in simulated surgical environments.

This article was submitted and is published as a preprint:

Rosenkranz, M., Uslar, V.N., Weye, D. & Bleichner, M. G. (2025, preprint). The effect of task demand on EEG responses to irrelevant sound and speech in simulated surgical environments. *bioRxiv*. https://doi.org/10.1101/2025.02.13.638036

Author contributions: MR and MGB conceptualized the experiment. MR performed the data acquisition and analyzed the data. MR wrote the manuscript to which MGB, VNU, and DW contributed with critical revisions. All authors reviewed the manuscript.

4.1 Abstract

Complex soundscapes in high-stakes environments, such as the operating room (OR), are characterized by a variety of overlapping auditory stimuli and present significant challenges for personnel, particularly during periods of high demand. This study investigates how task demand and an OR soundscape including irrelevant speech, influence perceived workload, surgical performance, and auditory processing in a simulated surgical environment, using mobile electroencephalography (EEG). Participants performed two simulated surgical tasks, namely peg transfer and suturing, representing a low-demand and high-demand task, respectively. The tasks were performed under two sound conditions: An OR soundscape was presented with irrelevant speech or alone. Neural responses to transient and continuous auditory stimuli were analyzed using event-related potentials (ERPs) and temporal response functions (TRFs), respectively. Results showed that irrelevant speech increased self-reported workload and distraction. EEG analyses revealed reduced neural responses to transient sounds and irrelevant speech under high task demand, reflecting early-stage sensory filtering of auditory distractions. Notably, an inverse relationship was observed between neural responses to speech and self-reported workload, indicating that the speech responses may serve as a marker for perceived workload. Overall, this study demonstrates the

potential of EEG to assess irrelevant sound processing in realistic work-like settings and highlights the critical role of task demand in modulating neural responses and self-reported workload to soundscapes.

4.2 Introduction

Certain professions require a high degree of skill and precision but must be performed in environments where distractions are inevitable and mistakes can have fatal consequences. Surgery, for example, is inherently demanding, requiring precise handling of instruments, advanced technological skills, and high levels of concentration over a long duration. In such professions, cognitive resources must be allocated to incoming information across multiple sensory modalities (Wickens, 2008). The environment is characterized by a complex soundscape, comprising a multitude of concurrent sounds, including communication, alarms, monitoring sounds, and instrument usage. The individual team member must distinguish between relevant and irrelevant, potentially distracting, information within this complex soundscape. Thus, distraction represents a cognitive challenge in the operating room (OR), with a notable impact on personnel well-being (Kern et al., 2019) and has been linked to elevated stress levels and an increased likelihood of errors (Mentis et al., 2016).

Although a variety of sounds can be perceived as distracting (Gülşen et al., 2021; Tsiou et al., 2008; Weigl et al., 2015), irrelevant speech, that is, speech unrelated to the procedure, is particularly perceived as distracting (Healey et al., 2007; Tsiou et al., 2008; van Harten et al., 2021) and increases perceived workload (Weigl et al., 2015; Wheelock et al., 2015). Although speech is frequently identified as a distractor in the OR, there is mixed evidence regarding the impact of the OR soundscape, including speech, on surgical performance. Some studies showed performance reductions when the soundscape was compared with silence (e.g., Pluyter et al., 2010; Siu et al., 2010) or when irrelevant speech is the sole distractor (Czerwiec et al., 2024). Yet, a realistic OR environment includes multiple overlapping sounds, with silence being a rare condition and irrelevant speech only one of several potential auditory distractors (Gülşen et al., 2021). This highlights the need to study how the combination of irrelevant speech and other overlapping sounds in the OR soundscape influence the individual and affect surgical performance under realistic conditions.

Instead of comparing an OR soundscape or speech with silence, auditory distraction should be investigated in the context of varying task demands. Subjective reports from medical personnel indicate that irrelevant speech is perceived as particularly distracting during phases of high task demand compared to phases of low task demand (Persoon et al., 2011; van Harten et al., 2021). This suggests that surgical task demand modulates the impact of irrelevant speech distractions. Understanding the contribution of distractors like irrelevant speech and their interaction with task demands is therefore crucial for optimizing performance and improving the work environment.

A comprehensive assessment of the interaction between task demand and distraction in complex environments like the OR requires a combination of objective and subjective measures. Using only subjective measures (e.g., self-reports), it can be challenging to separate the specific contribution of speech from the overall impact of the task demand, and other sounds in the OR environment (Dias, Ngo-Howard, Boskovski, Zenati, & Yule, 2018). Similarly, measures of performance may not capture situations where individuals find the soundscape distracting, even if their performance is unaffected (Rosenkranz, Haupt, Jaeger, Uslar, & Bleichner, 2024). For example, surgeons may experience increased workload in order to maintain a high level of performance within a distracting environment. To complement self-reports and performance measures, electroencephalography (EEG) provides another measurement. Mobile EEG is increasingly employed in the study of work environments (Wascher et al., 2021) and provides reliable responses to complex soundscapes and speech while a task is being performed (Herrmann, 2024; Rosenkranz et al., 2023, 2024; Xie et al., 2023). Thus, EEG can be related to the different potential auditory distractors like irrelevant speech and the OR soundscape and assess how the processing of each distractor varies with task demand.

To investigate the neural processing of distinct auditory stimuli within the OR soundscape, we employed two complementary EEG analysis approaches. First, we computed event-related potentials (ERPs) which are well-established for studying the processing of transient sounds in relation to task demands (e.g., Wascher et al., 2021). This makes ERPs particularly well-suited to examine how surgical task demand influences the processing of task-irrelevant auditory stimuli. To account for the continuous aspects of the OR soundscape, we also computed temporal response functions (TRFs), a robust tool for capturing neural responses to continuous auditory stimuli, including speech and concurrent sounds (Crosse et al., 2016; Rosenkranz et al., 2023, 2024) By using both ERPs and TRFs, we aim to gain a more comprehensive understanding of how discrete and continuous auditory stimuli are processed and how task demand shapes the neural responses in a complex work-like environment.

The ability to filter out irrelevant auditory information is essential for maintaining focus on a task. Sensory gating, a neural mechanism thought to help suppress responses to repetitive, irrelevant stimuli, may play a key role in this process (Lijffijt et al., 2009). To investigate sensory gating without interfering with the surgical task, we employed the paired-click paradigm. This approach contrasts the neural response to an initial click to a repeated second click. The reduction in response from the first to the second click reflects the strength of sensory gating. Studies suggest that greater cognitive engagement in a task increases the response reduction in the paired-click paradigm, indicating more effective inhibition of irrelevant stimuli (Lijffijt et al., 2009). Similarly, an increase in task demand can suppress the processing of irrelevant auditory stimuli when the relevant stimuli are of a different sensory modality (Molloy, Lavie, & Chait, 2019; Sörqvist et al., 2016; Sörqvist, Stenfelt, & Rönnberg, 2012). Sensory gating effects have been shown to remain robust even when individuals are engaged in cognitive tasks or exposed to background noise (Hölle &

Bleichner, 2023b; Major et al., 2020), making this method interesting for studying auditory filtering under real-world conditions. Based on these findings, we hypothesized that high task demand would enhance sensory gating, resulting in a larger difference between ERPs in responses to the first and second clicks, compared to low task demand.

Besides examining ERPs in response to the paired-click paradigm, we investigated whether the reduction in ERP amplitudes under high compared to low task demand (Molloy et al., 2019; SanMiguel et al., 2008; Sörqvist et al., 2016) extends to neural responses to the entire soundscape. Unlike the repetitive stimuli typically used in ERP studies, natural soundscapes consist of distinct and concurrent, as well as, overlapping auditory events. To better capture how such soundscapes are processed, we computed a general neural response using TRFs. This approach allowed us to assess whether the reduction in ERP responses under high task demand could also be observed when analyzing the continuous OR soundscape.

For speech processing, we expected it to be modulated by task demand as well, though the direction of this effect is uncertain. While prior studies indicate that processing of irrelevant non-speech sounds decreases with increasing task demand (SanMiguel et al., 2008; Sörqvist et al., 2016), speech has a higher potential to distract than non-speech sounds (Szalma & Hancock, 2011). This makes speech particularly disruptive for OR personnel, especially during phases of high task demand (van Harten et al., 2021; Weigl et al., 2015; Widmer et al., 2018). Given that distracting non-speech sounds enhance the neural response (Huang & Elhilali, 2020), we considered the possibility that the neural response to speech could be enhanced during high-demand phases, even as non-speech sounds are filtered out.

In summary, this study examined how an OR soundscape including irrelevant speech and varying task demands interact to influence self-reports, surgical task performance, and auditory processing, as measured by EEG. We hypothesized that higher task demand will lead to an increase in perceived workload, while the presence of irrelevant speech will increase distraction ratings. Additionally, we explored the interaction between task demand and speech presence to assess their combined influence on self-reported workload and distraction. As workload is a rather general measure of surgical task demand, we also explored how the tasks and speech affect specific aspects of self-reported demand (Wilson et al., 2011). We further explored the effect of irrelevant speech on surgical task performance. Regarding neural responses, we expected that higher task demand will result in an increased sensory gating, as measured by ERPs. For TRF responses, we hypothesized that increased task demand will lead to a lower neural response to an OR playback. Furthermore, we explored the relationship between task demand and speech processing, as well as the association between neural responses to the OR playback and speech with self-reported workload.

4.3 Method

This study involving human participants was reviewed and approved by Medizinische Ethikkommission, Carl von Ossietzky Universität Oldenburg, Oldenburg (2021-031). The participants provided their written informed consent prior to participating in this study. All participants received monetary reimbursement.

4.3.1 Participants

25 participants were recruited through an online announcement on the University board (age range: 19-34; mean age = 24.84; 14 female; 10 medical students). Eligibility criteria included: self-reported normal hearing, normal or corrected vision, no psychological or neurological condition, right-handedness, and no experience with surgery or surgical simulations. In total five participant were excluded from all analyses involving EEG, due to the following reasons: no data were present due to a recording error (N=1); package loss during recording resulted in timing problems (N=2); connector problems resulted in artifactual data (N=2). Those participants were still included in the self-report and performance analyses.

4.3.2 Paradigm

We employed a within-subject design, featuring a 2 (task: easy vs difficult) x 2 (sound: speech present vs speech absent) factorial structure, where each condition was repeated over three blocks, resulting in a total of 12 blocks (Figure 4.1). The participants were required to complete the surgical tasks peg transfer and suturing. These tasks were selected as they have been demonstrated to elicit either low or high workload, respectively (Lim et al., 2023; Scerbo, Britt, & Stefanidis, 2017). During all blocks, participants were presented with a task-irrelevant soundscape. In all blocks the soundscape consisted of sounds from an actual OR (i.e., OR playback), and click sounds. Additionally, in half of all blocks, speech was presented. Each block lasted up to six minutes, with the end of the block signaled by the soundscape fading out, indicating that participants should stop the surgical task. Participants were instructed that all auditory stimuli were irrelevant and could be ignored.

Auditory stimuli

We generated nine auditory stimulus versions, each lasting six minutes, which were consistent across all participants (Figure 4.1b). Of these, three versions featured each a different OR playback with click pairs distributed throughout the playback duration and were used for the speech-absent conditions. The remaining six versions were used for the speech-present conditions, created by pairing each speech-absent version with one of three speakers, with each speaker narrating two different stories. The OR playback and

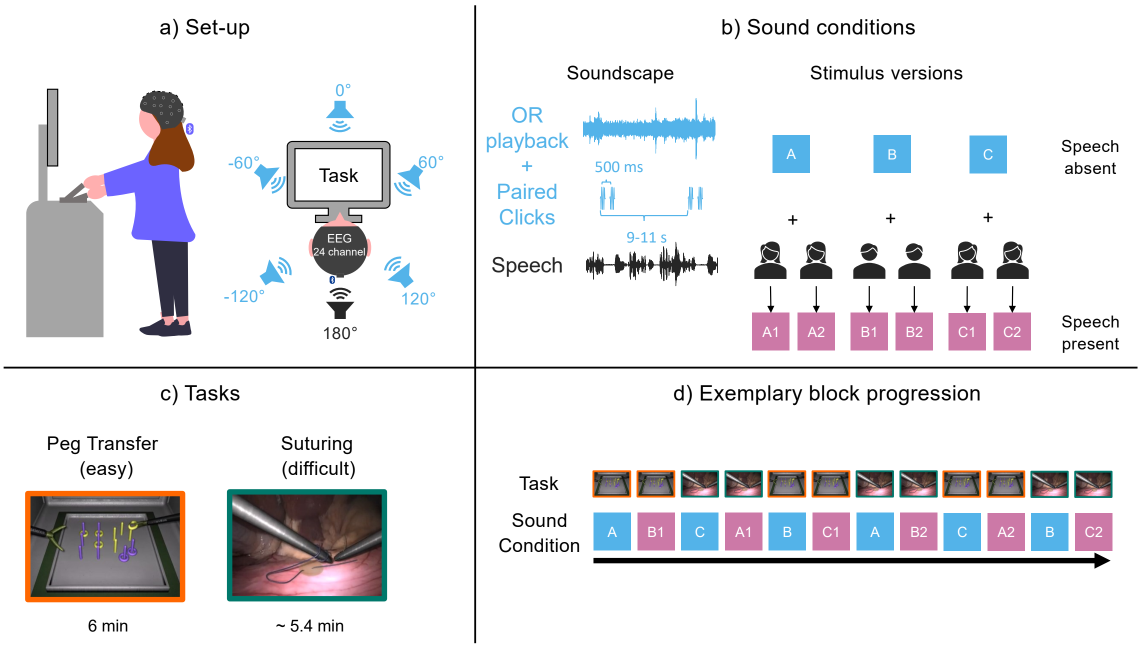


Figure 4.1: a) Participants performed surgical training tasks while standing in front of a surgical simulator, equipped with a 24-channel mobile EEG cap. A soundscape was presented through a loudspeaker array positioned around them. From five loudspeakers, marked in blue, an OR playback and paired clicks were presented. From one loudspeaker, marked in black, speech was presented. b) All sound conditions (i.e., speech-absent and speech-present) included the OR playback and paired clicks. Additionally, stories from three speaker (two stories per speaker) were presented, with each speaker paired with one of the three OR playbacks. This resulted in six stimuli for the speech-present condition. c) Participants completed two tasks of varying difficulty: peg transfer (easy) and suturing (difficult). The peg transfer task was performed until the end of a block (after six minutes). The suturing task was performed until it was finished (which was on average after 5.4 min), but no longer than six minutes. d) Example block progression for one participant. Each task was presented in two consecutive blocks, alternating between speech-present and speech-absent conditions. The starting task and sound condition was counterbalanced across participants. Additionally, while the overall order of sound conditions was fixed, the starting sound condition was rotated between participants.

paired clicks were presented through five loudspeakers positioned around the participant (0°, 60°, 120°, -120°, -60°), while speech was presented through a single loudspeaker positioned behind the participant (180°, Figure 4.1a). Auditory stimuli were presented using Psychtoolbox 3 (v3.0.17, Kleiner et al., 2007). For each stimulus type, a sound marker was generated using the lab streaming layer library (v1.14, Kothe et al., 2024).

OR playback The three OR playbacks were extracted from a recording during a visceral surgery using a field recorder which was positioned close to the surgery table at the University Hospital Oldenburg (Rennies et al., 2023). The recording contains a variety of sounds, such as ventilation noise, beeps from monitoring devices, instrument clatter, and instrument sounds. Intelligible speech was removed after the recording for privacy reasons, however, unintelligible muttering and non-vocal sounds such as coughing were preserved.

Paired clicks We presented pairs of clicks, with an interval of 500 ms between clicks and an interval of 9-11 sec between click pairs. Each click pair consisted of two identical clicks (1000 Hz, 4 ms duration, 1 ms onset and offset ramps). In total, 35 pairs were presented per block. To ensure that energetic masking influences of the clicks was similar across participants, all pairs were presented at fixed moments in the OR playback.

Speech stimuli The speech stimuli were chosen from a database containing German speakers, who were telling stories about self-selected content (Daeglau et al., 2023), and have been shown to provide measurable EEG responses (Daeglau et al., 2025; Wiedenmann et al., 2023). The natural speech included speech pauses and filler words which increased the ecological validity of our approach. Three speakers were chosen, each telling two stories. To control for differences in loudness, the speech stimuli were matched to have the same root-mean-square (RMS) value.

Surgical tasks

For the surgical task, the LabSim® (Surgical Science, Sweden) simulator was used. The simulator includes the surgical training tasks peg transfer and suturing which were chosen as they differ in difficulty (Figure 4.1c; Lim et al., 2023; Scerbo et al., 2017), required bi-manual control, and lasted at least three minutes for inexperienced individuals, ensuring that sufficient data could be collected. This minimum duration was an approximation based on observations from pilot data. As the tasks varied in their goal and procedural steps, the performance measures were different between tasks. Peg transfer included the performance measures ’number of transfers’ and ’number of drops’, while suturing included the performance measures ’duration’ and ’damage’. Although participants were not provided with any feedback regarding their performance during the task, they were instructed as to the performance measures that we investigated. Participants always completed two consecutive blocks of one task (e.g., peg transfer) and then switched to two consecutive blocks of the other task (e.g., suturing).

Peg Transfer In the peg transfer task, participants were required to transfer rings between two pairs of pegs, including a switch between grasping instruments for each transfer. This task was defined as easy, as it involved only few and repetitive procedural steps. Participants were instructed to complete as many transfers as possible within a block with minimal ring drops. The task automatically ended after six minutes. The number of transfers and drops were used as the performance measures; these values were not directly provided by the LabSim software but could be compute based on the number of grasps, average drops, and average transfers.

Suturing In the suturing task, participants were required to drive a needle through tissue and tie two knots in the suture thread using the provided instruments. This task was more difficult than the peg transfer, as it included several procedural steps, and required a higher degree of dexterity. Once the second knot was tied, the task ended and the soundscape stopped. The performance measure were task duration and damage. Damage was defined as the number of times the tissue was touched and the amount of pressure applied to the needle after it was driven through the tissue. This resulted in a damage score from 0 to 100, with a higher score indicating less damage.

Counterbalancing of blocks

To counterbalance the order of tasks, half of the participants started with peg transfer, while the other half started with suturing. The experiment followed a fixed order of stimulus conditions where speech absence and presence alternated with each block (Figure 4.1d). To counterbalance the sequence of stimulus conditions across participants, participants began at a different starting point within this fixed order and then continued sequentially. Consequently, the order was repeated after the 12th and 24th participant. We further ensured that one story of each speaker was presented during peg transfer and the other story of each speaker during suturing.

Procedure

After arrival, participants practiced the use of the lab simulator for 45 min with the following procedure. To get acquainted to the simulator and the instruments, two simple training tasks (i.e., instrument navigation and grasping) were repeated twice each. This was followed by two blocks of peg transfer and two blocks of suturing. During the first block of each task, the experimenter provided instructions and guidance. During the second block of each task, the experimenter left the room and an OR playback was presented. The OR playback was not used during the experimental blocks. Participants were always allowed to ask questions and watch short instruction videos, provided by the manufacturer of the simulator. After the training, the EEG cap was fitted and participants performed resting measurements (i.e., two minutes of eyes open, two minutes of eyes closed, and listening to a sequence of 20 beeps). After these, participants performed the 12 experimental blocks. After each block participants completed the SURG-TLX (Wilson et al., 2011) thereby providing a self-reported workload measurement. The SURG-TLX contains six items related to the different aspects of surgical demand (mental demand, physical demand, temporal demand, complexity of procedure, stress, and distraction). The items were rated on a visual analogue scale with scores ranging from 0 (low) to 20 (high). At the end of the experiments, participants answered 12 questions regarding the content of the speech stimuli (e.g., Was the topic of one of the stories a bike?). Six questions were related to the content of the stimuli - one question per stimulus - while six questions were unrelated. The participants had to indicate whether they perceived the content or not. An inspection of the speech content questions showed that 86,67% of the questions were answered correctly. Thus, participants could discriminate between speech-related and speech-unrelated questions.

4.3.3 EEG data acquisition

Participants were asked to wash their hair at the day of recording prior to the experiment. EEG data were recorded using a wireless amplifier (SMARTING, mBrainTrain, Belgrade, Serbia) attached to the back of a 24-channel EEG cap (EasyCap GmbH, Hersching, Germany) with Ag/AgCl passive electrodes and the reference and ground electrode at position FCz and AFz, respectively. The data were recorded using a sampling rate of 500 Hz, and transmitted via Bluetooth from the amplifier to a Bluetooth dongle (BlueSoleil) that was plugged into a computer (DELL Precision 3630).
After fitting the cap, the skin at each electrode site was cleaned using 70% alcohol. Skin conductance between the scalp and electrodes was increased using abrasive gel (Abralyt HiCl, Easycap GmbH, Germany). Impedance were kept below 10 kΩ at the beginning of the recording.

The transmitted EEG data and sound marker were recorded in the Lab Recorder software [1] and saved as .xdf files. The same computer was used for data recording and sound presentation. Due to technical reasons, a constant delay between the marker indicating sound onset and the actual sound presentation was measured in advance. To account for this delay, the marker was adjusted offline by shifting it 40 ms.

4.3.4 EEG preprocessing

The EEG data were analyzed using EEGLAB (v2022.1, Delorme & Makeig, 2004) in MATLAB R2020b (The MathWorks, Natick, MA, United States). As a first step, bad channels that were recognized during recording were removed, resulting in the removal of one channel for two participants. After bad channel removal, the data were cleaned from artifacts using infomax independent component analysis (ICA). To improve ICA, the data were first high-pass filtered (passband edge = 1 Hz[1]), low-pass filtered (passband edge = 40 Hz[2]), and resampled to 250 Hz. Then, the resting data and data of each block during which audio was presented, were cut into consecutive epochs of one second. To minimize artifacts from the start and end of the task the first and last five seconds of each block were excluded. Improbable epochs with a global (all channels) or local (single channel) threshold exceeding 5 standard deviations were automatically rejected using the *jointprob* function. ICA decomposition was applied to the remaining epochs. The resulting components were back-projected on the raw data. The raw data were then high-pass filtered (passband edge = 1 Hz[1]) and low-pass filtered (passband edge = 25 Hz[3]). The back-projected components were then classified using the EEGLAB toolbox *ICLabel* (Pion-Tonachini et al., 2019) with the 'lite' classifier which is better at detecting muscle artifacts than the default classifier (Klug & Gramann, 2021). Components belonging to the categories eye blink and

[1] https://github.com/labstreaminglayer/App-LabRecorder, v1.14
[1] filter order = 825, transition bandwidth = 1 Hz, cutoff frequency (-6dB) = 0.5 Hz
[2] filter order = 83.5, transition bandwidth = 10 Hz, cutoff frequency (-6dB) = 45 Hz
[3] filter order = 132.5, transition bandwidth = 6.25 Hz, cutoff frequency (-6dB) = 28.125 Hz

movement or muscle movement with 70% confidence or heart with 80% confidence were removed. Note, that the ICLabel classifier did not classify all components correctly because it was trained on stationary data with a larger electrode setup than ours. Therefore, we manually checked the components and made the following adjustments: Components indicating lateral eye-movement were not always correctly classified and individually removed. Furthermore, channel Tp9 and Tp10 contained noise from muscle movement. We observed this already in a previous experiment, where a surgical simulator was used (Rosenkranz et al., 2024). We assume that the bi-manual control of the simulator activates neck muscles, resulting in artifacts in electrodes that are close to the neck. As channel Tp9 and Tp10 are used for re-referencing, we removed components showing muscle activity in Tp9 and Tp10. Afterwards, previously rejected channels were interpolated using spherical interpolation. Lastly, channels were re-referenced to the linked mastoids (Tp9 and Tp10).

4.3.5 ERP analysis

We analyzed the neural response to the irrelevant soundscape using two different approaches. Event-related potentials (ERPs) were computed in response to the paired clicks, while a temporal response function (TRF) was used to analyze the neural response to the OR playback and speech. Since the irrelevant speech was presented separately to the OR playback, we computed separate TRFs for these.

We quantified ERP amplitudes using the following procedure: For each block we extracted epochs from -200 to 1000 ms relative to the onset of the first click and baseline-corrected the epochs from -200 to 0 ms. Epochs exceeding a threshold of 3 standard deviations globally (across all channels) or locally (within a single channel) were automatically rejected using the *jointprob* function. Data from channel FC1, FC2, Fz, and Cz were then averaged, as the auditory N1 and P2 ERP components are prominent at these channels (Crowley & Colrain, 2004; Näätänen & Picton, 1987). To extract the N1 and P2 components for both the first and second clicks, we first computed an average ERP across participants for each sound condition. It should be noted that the extraction of time windows was conducted for the two sound conditions separately to account for the acoustic differences between the two conditions. For each sound condition, we identified the N1 peak within the range of 80 to 140 ms. Amplitudes were then averaged within a $\pm$ 25 ms range around the peak, resulting in one N1 amplitude value per participant and task for the first click. The P2 component was identified similarly, with a peak search window of 150 to 250 ms and a $\pm$ 25 ms range around the peak. For the second click, we added 500 ms to the N1 and P2 time-window of the first click, and averaged across this time-window. We performed a peak-to-peak analysis by subtracting the N1 amplitude from the P2 amplitude for each click. The resulting difference score defined the response amplitude for each click, sound condition, task, and participant. The difference between the first and second click defined the gating value.

4.3.6 TRF analysis

Audio preprocessing

To relate the ongoing soundscape to the ongoing neural response, we extracted the envelope of the OR playback and speech. The OR playback included noise from running machines and ventilation, which produced an envelope with little variation. Such low variability in the envelope can lead to poor TRF estimation (Rosenkranz et al., 2023). To address this, we applied a Wiener filter implemented in MATLAB (Plapous et al., 2006; Scalart, 2023). For this, we first high-pass filtered each OR playback at 1 Hz[4]. We then estimated the power spectral density of the noise using the first second of each OR playback, as it was representative of the static noise in the OR playback. The noise estimate was then subtracted from the remaining signal. Afterwards, we extracted the envelope from the noise-reduced OR playbacks and raw speech using the *mTRFenvelope* function (Crosse et al., 2016, 2021). We resample all envelopes to 125 Hz.

Backward modeling

A backward modeling approach was utilized to analyze the response to each stimulus envelope separately using the mTRF toolbox (Crosse et al., 2016, 2021). In backward modeling, the neural response is used to reconstruct features of the stimulus. The procedure involved a nested cross-validation for each task and stimulus. As a first step, the EEG data were resampled to 125 Hz to match the sampling rate of the envelopes. The first and last five seconds of the EEG data and envelope of each block was removed. Each block was split in half, resulting in six segments per task. The mean duration of each segment was 175 and 162 seconds for the peg transfer and suturing task, respectively. Each segment served once as a test set while the remaining segments served as training sets, resulting in six folds. For each fold, the following procedure was employed: The training segments were cross-validated to determine the optimal regularization parameter (i.e., lambda) using the function *mTRFcrossval*. The optimal lambda was searched for in the range of $10.\hat{}[-3 : 1 : 7]$. A time-window from 0-300 ms was chosen, as irrelevant stimulus processing typically takes place in earlier time-windows (Hausfeld, Riecke, & Formisano, 2018; Hausfeld, Shiell, Formisano, & Riecke, 2021). The lambda corresponding to the maximum correlation coefficient was selected for model training. The model was then trained using the *mTRFtrain* function and the optimal lambda. The trained model was then applied to the test set using the *mTRFpredict* function, which correlated the actual and predicted stimulus envelope. This resulted in one Pearson correlation value (i.e., prediction accuracy) per fold which were averaged across folds, resulting in one correlation value per participant, task, and stimulus. In all conditions, we analyzed the neural response to the OR playback. Additionally, in the speech-present condition, we applied the same analysis procedure to the speech stimuli. For this, we computed TRFs using the combined speech

[4] filter order = 1000, transition bandwidth = 0.5 Hz, cutoff frequency (-6dB) = 0.00004 Hz

material, disregarding differences in content, speaker sex, and speech characteristics, such as word frequency.

We also investigated whether the correlation between the actual and reconstructed stimulus envelope was above chance. For this, a permutation test was conducted using the *mTRFpermute* function. This involved shuffling the response data and recalculating the correlation to generate a distribution of correlation coefficients under the null hypothesis, representing chance-level performance. In total 17 permutations were performed per fold, resulting in 102 permutations per participant, task, and stimulus. The permutation test of each fold used the optimal lambda of the respective fold. The chance-level was defined as the 95th percentile across all permutations of a stimulus.

4.3.7 Statistical analyses

All statistical analyses were performed in R Studio (v4.2.1). For most outcome measures, we fitted a series of linear mixed models using the R packages lmer4 (v1.1-30) (Bates, Mächler, Bolker, & Walker, 2015). We subsequently added fixed then random effects and evaluated the improvement in models fit. The baseline model always included the random intercept of the participant. For most models, the predictors of *task* and/or *sound* were added. *Task* contained two categories, peg transfer and suturing, which were coded 0 and 1, respectively. *Sound* contained two categories, speech absent and speech present, which were coded 0 and 1, respectively. The best fitting model was determined using likelihood-ratio testing. We report results from the likelihood-ratio test comparing a model with a fixed or random effect to a model without the effect. We further report for the fixed effects the b value and standard error (SE) of the best fitting model. The model comparisons for all computed models can be found in the supplementary materials (section 'Model comparisons').

Self-reported workload

For self-reported workload and distraction we used two outcome measures from the SURG-TLX. First, the total SURG-TLX score defined the workload and was calculated by averaging the scores for all items. For each task and sound condition, the mean score was calculated across blocks. Secondly, the score for the item distraction was calculated by averaging across blocks. This resulted in 2 (task: peg transfer vs suturing) x 2 (sound: speech present vs. speech absent) workload and distraction scores for each participant. We expected the total score to change between the tasks, and explored the effect of the sound condition. Therefore, we iteratively added first *task*, second *sound*, and then their interaction as fixed effects. After fitting the fixed effects, we fitted the random effects by adding *task* and *sound* as random slopes.

For the distraction item we expected it to differ mainly between sound conditions, thus we first added *sound*, second *task*, and then their interaction. After fitting the fixed effects, we fitted the random effects by adding *sound* and *task* as random slopes.

While the total score and distraction item were our main outcome measure, we also explored how *task* and *sound* affect the different aspects of surgical demand. Therefore, we followed the same procedure as for the total score also for the other items of the SURG-TLX (i.e., mental demand, physical demand, temporal demand, complexity, and stress).

Surgical task performance

We computed separate linear mixed models for the outcome measures of the peg transfer and suturing tasks. As the outcome parameters were different for each task, we did not investigate whether performance differed across tasks. Instead, we focused on whether the presence of speech influenced task performance. For each outcome parameter, we investigated the effect of adding the fixed and random slope of *sound*. For the outcome measure 'duration' the data were skewed, as most participants did not complete the task within the 6 minutes. Therefore, we computed a beta mixed model (Verkuilen & Smithson, 2012) using the R package *glmmTMB* (v.1.1.7). For this, we normalized all values between 0 and 1 with 1 indicating that the entire duration was used. As beta models do not allow values to be exactly 1 we transformed boundary values by subtracting 0.002. Otherwise, the same method as with the other outcome parameters was applied.

Sensory gating

We use the paired-click paradigm to assess the amount of sensory gating for each task. We first evaluated whether a gating effect was present. We did this separately for the sound conditions, as they were acoustically different, thus different responses could be expected. For each sound condition we evaluated whether the response to the first and second click were different, in other words, whether gating was present. We computed a linear mixed model for each sound condition using the response amplitudes to the clicks as an outcome measure. The baseline model included *participant* as a random intercept and *position* (coded 0 and 1 for the response to the first and second click, respectively) was subsequently added as a fixed effect and random slope.

We then checked, whether gating differed between tasks. The gating value was obtained by subtracting the amplitude of the first click from that of the second click. This gating value, calculated for each participant and task, reflects the strength of the sensory gating, with a larger value indicating a larger gating, in other words, better suppression of the second click after hearing the first one. For each sound condition, we computed the baseline model including *participant* as a random intercept and subsequently added *task* as a fixed effect and random slope.

Continuous stimuli

For the response to the continuous stimuli, three prediction accuracies were computed. For the speech-absent condition, only the playback stimulus was modeled, resulting in one prediction accuracy. For the speech-present condition, the playback and speech were modeled individually, resulting in two prediction accuracies. Each prediction accuracy was modeled using the baseline model *participant* as a random intercept and subsequently adding *task* as a fixed effect and random slope.

Exploratory analyses

We investigated the relationship between self-report and neural measurements. To do this, we used the prediction accuracies (i.e., *r*) as continuous predictors for the total workload score. We calculated the average workload score for both the blocks where speech was absent or present. In both the speech-absent and speech-present conditions, the prediction accuracy of the OR playback was used to predict the total workload for each respective condition. In addition, in the speech-present condition, the prediction accuracy of the speech stimulus was also used to predict the total workload for that condition. To receive meaningful b estimates we centered the prediction accuracies for each stimulus. To account for the effect of task, *task* was also included as a predictor. Thus, the baseline model included the random effect *participant* and fixed effect *task*. We subsequently added the fixed effect *r* and the interaction between *r* and *task*. This was done for each stimulus separately.

4.4 Results

4.4.1 Self-reported workload

Total score Most participants reported a higher overall workload during the suturing compared to peg transfer task (Figure 4.2a). The best fitting model included the main effects *task* ($\chi^2(1) = 85.86, p < .001, b = 4.58, SE = 0.5$) and *sound* ($\chi^2(1) = 12.01, p < .001, b = 1.2, SE = 0.27$), but no interaction ($\chi^2(1) = 0.0001, p = .993$). When allowing the effect of *task* and *sound* to vary across participants, the model fit further improved (*task*: $\chi^2(1) = 24.68, p < .001$; *sound*: $\chi^2(1) = 10.26, p = .016$). This indicates that participants experienced higher self-reported workload during the suturing task compared to the peg transfer task, a higher workload when speech was present compared to when speech was absent, and that the strength of both effects varied between participants.

Distraction score The best fitting model included the main effects *sound* ($\chi^2(1) = 60.30, p < .001, b = 4.79, SE = 0.69$) and *task* ($\chi^2(1) = 6.78, p = .009, b = 1.27, SE =$

0.33), but no interaction ($\chi^2(1) = 0.66, p = .415$). Allowing the effect of *sound*, but not of *task*, to vary across participants further improved the model fit (*sound*: $\chi^2(2) = 19.37, p < .001$, *task*: $\chi^2(3) = 2.30, p = .513$). This indicates that participants experienced higher levels of distraction during the suturing task compared to the peg transfer task, more distraction in the presence of speech compared to its absence, and that the effect of *sound* varied between participants.

Exploratory analyses of individual SURG-TLX items We investigated the remaining SURG-TLX items individually (see supplementary figure S4.1) and listed the results in the supplementary table S4.1. To summarize, we subsequently added the fixed effects, *task*, *sound*, and their interaction, and the random slopes. Adding *task* as fixed effect as well as random slope significantly improved model fit for the items mental demands, physical demands, temporal demands, and complexity of procedure. For the same items, adding *sound* or the interaction between *task* and *sound* did not improve the model fit significantly. For the item situational stress, the best fitting model included the main effects for *task* and *sound* and random slopes for both effects. To summarize, while all items showed higher scores for the suturing compared to peg transfer task, only the distraction and stress item showed also higher scores for the speech-present compared to speech-absent condition.

4.4.2 Surgical task performance

For the peg transfer task, the outcome parameters were the number of transfers and the number of drops. For the suturing task, the outcome parameters were the time required to complete the task and the extent of damage. As shown in Figure 4.2 c)-f), adding *sound* as predictor did not change model performance for any outcome measure (Transfers: $\chi^2(1) = 0.021, p = 0.886$; Drops: $\chi^2(1) = 0.127, p = 0.7215$; Time: $\chi^2(1) = 0.22, p = 0.64$; Damage: $\chi^2(1) = 2.267, p = 0.132$). This indicates that the presence of speech did not change the measured task performance during either surgical task. Furthermore, there was a high variability in performance between participant. This was probably the result of our participant pool, that had little laparoscopic experience.

4.4.3 Sensory gating

To investigate the presence of sensory gating, we computed whether the ERP amplitude changed from the first to the second click. We observed a gating effect in the speech-present and speech-absent condition. For both sound conditions, the model fit improved when adding *position* as a predictor (Figure 4.3b & e; speech-absent condition: $\chi^2(1) = 29.07, p < .001, b = -1.1, SE = 0.21$; speech-present condition: $\chi^2(1) = 7.81, p = .005, b = -0.6, SE = 0.21$). Allowing *position* to vary across participants improved the model fit for the speech-absent condition ($\chi^2(1) = 11.02, p = .004$) but not for the speech-present condition ($\chi^2(1) = 2.67, p = .263$). Figure 4.3a) and d) shows that the ERP morphologies

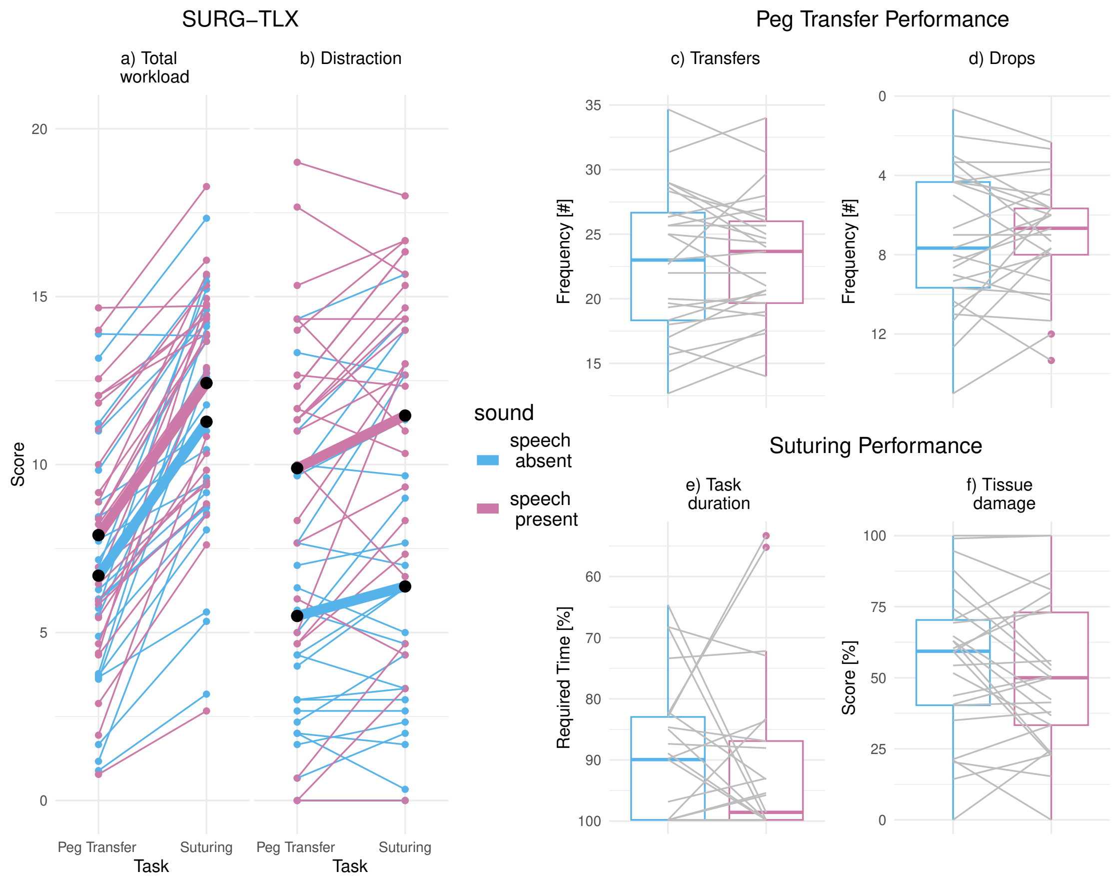


Figure 4.2: Self-report and performance measures for each task and sound condition. The tasks had two difficulty levels, the easy peg transfer task and the difficult suturing task. Self-reports were derived from the SURG-TLX. a) The total workload score (task & sound: $p < .001$). b) The score on the distraction item (task & sound: $p < .01$). Thin lines represent individual data for each sound condition, thick lines the average across participants. To evaluate the effect of speech on surgical task performance we used for the peg transfer task c) the number of transferred ring, and d) the number of dropped rings. For the suturing task we used e) the amount of the total time that was required (i.e., 0-6 min, Mean = 5.4 min), and f) the efficacy of handling the tissue. For all performance plots, a high y-axis value indicates good performance, and a low value bad performance. The gray lines show individual participants. None of the performance measures was significantly affected by the sound condition.

between the sound conditions are different, likely due to the different acoustics in each sound condition (i.e. the presence or absence of speech).

We then investigated whether the sensory gating strength changed between tasks. For the speech-absent condition, adding *task* as a predictor significantly improved model fit (Figure 4.3c; $\chi^2(1) = 6.75, p = .009, b = 0.791, SE = 0.287$), but adding *task* as random slope led to an unidentifiable model. For the speech-present condition, adding *task* as a predictor did not improve model fit, compared to the baseline model (Figure 4.3f; $\chi^2(1) = 2.4, p = .13$). To summarize, sensory gating was stronger in the suturing compared to peg transfer task in the speech-absent condition.

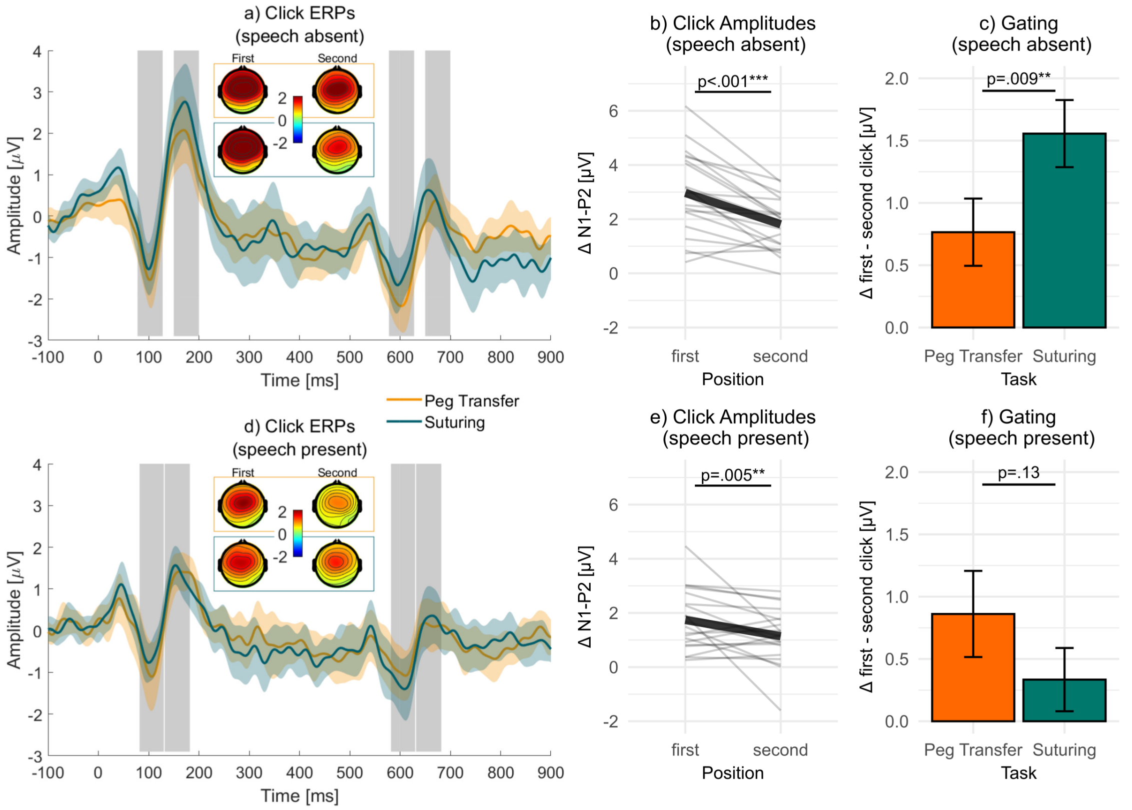


Figure 4.3: ERPs in response to the paired-click paradigm. The top and bottom row show data for the speech-absent and speech-present condition, respectively. The first column (a&d) shows the ERP time-course and topographies in response to the clicks, which were presented at 0 and 500 ms. The time-course shows the averaged data across participants and channels FC1, FC2, Fz, and Cz (solid line), along with the confidence interval (shaded area) for each task. The gray areas highlight the N1 and P2 time windows. The topographies show the peak-to-peak difference from the averaged amplitudes in the N1 and P2 time-window for the first and second click. The second column (b&e) shows the averaged amplitude for the first and second click. The thick line shows the average across participants and thin lines individual participants. The third column (c&f) shows the strength of gating, that is the difference between the response to the first and second click for each task ($\pm 1SE$).

4.4.4 Responses to continuous stimuli

Figure 4.4 (top row) shows that all models performed above chance. For the OR playback, adding *task* as a predictor neither increased model performance when speech was absent ($\chi^2(1) = 1.56, p = 0.135$) nor when speech was present ($\chi^2(1) = 0.625, p = 0.54$). For the speech stimulus, adding *task* as a predictor improved model performance ($\chi^2(1) = 11.723, p < 0.001, b = -0.015, SE = 0.004$), but adding *task* as random slope led to an unidentifiable model. This indicates that the prediction accuracy for speech was higher for the peg transfer than for the suturing task.

4.4.5 Relationship between self-report and neural responses

We explored the relationship between the self-report workload measures and neural responses to the continuous stimuli (Figure 4.4 bottom row). Adding *r* as predictor did not change model performance for the OR playback when speech was absent ($\chi^2(1) = 0.803, p = 0.37$) nor when speech was present ($\chi^2(1) = 0.738, p = 0.39$). However, adding *r* as predictor significantly improved model performance for speech ($\chi^2(1) = 4.687, p = 0.03$, $b = -1.39$, $SE = 0.64$), indicating that a lower neural response to speech was associated with larger perceived workload. Adding the interaction of *task* and *r* did not improve the model further ($\chi^2(1) = 0.041, p = 0.84$), and adding *task* or *r* as a random factor lead to unidentifiable models.

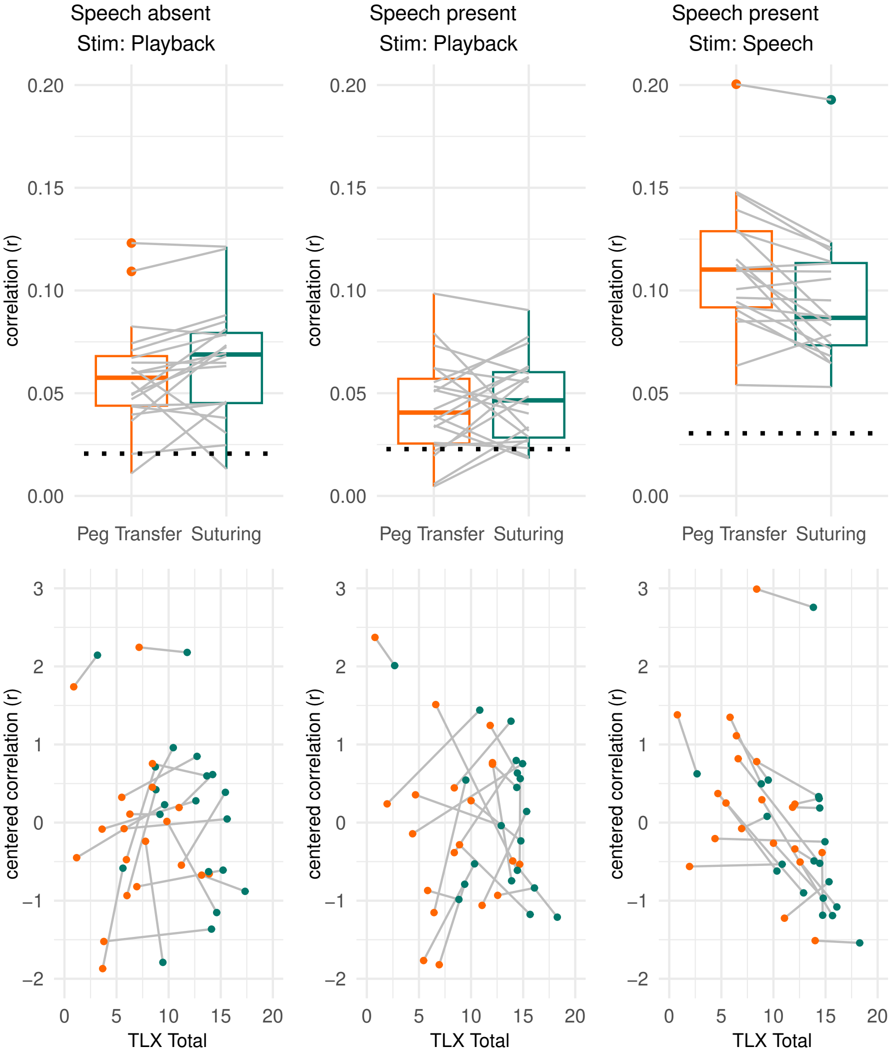


Figure 4.4: The plots show the effect of task on the correlation between the predicted and reconstructed stimulus (top row) and the effect of task and centered correlation on the total workload score (bottom row). Each column represents reconstruction accuracies for one of the three stimuli (i.e., Stim). Grey lines show data for individual participants. For the top row, the chance level is represented by the black dotted line. For the bottom row, we predicted the workload score using the centered correlation, but plotted the workload score on the x-axis and centered correlation on the y-axis, to visually match the top row.

4.5 Discussion

While speech distraction in the OR is an often reported problem, only few studies investigated this experimentally. Therefore, we studied how a soundscape, consisting of an OR playback, paired clicks, and irrelevant speech is perceived and processed during the performance of an easy and difficult surgical task, namely peg transfer and suturing, respectively. To understand how the soundscape is processed and influences the individual, we employed self-report, performance, and neurophysiological measurements.

As expected participants reported higher workload for the suturing task compared to the peg transfer task. This aligns with previous research showing that suturing (i.e., the difficult task) to be more demanding than peg transfer (i.e., the easy task), thus resulting in higher workload (Lim et al., 2023; Scerbo et al., 2017). The presence of speech increased the perceived workload as reflected in SURG-TLX items *distraction* and *stress*. Speech likely introduced a salient distraction that was difficult to ignore. This parallels findings from the actual OR where disturbance due to irrelevant speech correlates most strongly with these two items (Weigl et al., 2015). This consistency highlights the realism of our setup in simulating speech distraction within the OR environment, and shows that irrelevant speech increases the overall perceived workload by increasing distraction and stress.

Participants reported feeling more distracted by their environment during the difficult task. This replicates a finding from our previous study where participants perceived the soundscape as distracting primarily when task demand was high (Rosenkranz et al., 2024). This effect is also consistent with reports from OR personnel, who often cite noise as particularly distracting during high-workload phases (van Harten et al., 2021). To address this, noise interventions should prioritize reducing unnecessary noise, especially during periods identified by surgeons as highly demanding. As potential solutions, one could consider either implementing a system to signal that the room should remain quiet, such as a traffic light system (e.g., Engelmann et al., 2014) or enabling surgeons to reduce potentially unavoidable, task-irrelevant sounds, by using hearing devices (e.g., Leitsmann et al., 2021; Rennies et al., 2023).

While participants' self-reports indicate more distraction when speech was present, their surgical task performance remained unaffected by speech. Similarly, noise reduction interventions in the OR may increase subjective well-being, without necessarily impacting patient outcomes (Engelmann et al., 2014; Leitsmann et al., 2021). We propose two potential explanations for this discrepancy. First, the maintenance of performance in the presence of task-irrelevant speech comes at the cost of increased workload. This compensatory workload may have long-term implications for surgeons' well-being, even if patients are not immediately affected (Ayas et al., 2022). Second, performance measurements may not be sensitive enough to detect subtle differences in behavior. For instance, expert and novice surgeons can achieve similar performance on simple surgical tasks, even when distracted (Hsu et al., 2008). Furthermore, the evidence regarding the effects of noise on surgical performance are rather mixed (Mentis et al., 2016), which may be caused by the heterogeneity in performance measures across studies. This highlights the limitations of current performance metrics which may only decline under severely distracting conditions, for instance when several distractors are combined (Szafranski et al., 2009). Overall, our results suggest that while the measured performance may not suffer, irrelevant speech adds a mental strain that is reflected in self-reports rather than in surgical performance. This suggests that improving the acoustic environment can benefit surgeons by reducing perceived distraction and workload, contributing to their overall well-being.

As the recording of self-reports is limited to a single point in time, and performance may not accurately reflect changes in the processing of the soundscape, we utilized EEG to objectively examine responses to the soundscape. We employed two approaches, namely ERPs and TRFs to study responses to the transient clicks and continuous stimuli, respectively. With regard to ERPs, a gating effect was observed in both sound conditions, with participants showing reduced neural responses to the second click compared to the first. This replicates previous findings of sensory gating in complex environments (Hölle & Bleichner, 2023b; Major et al., 2020). Sensory gating is a neural mechanism that reflects early processing, specifically the filtering of irrelevant information (Lijffijt et al., 2009). Thus, our results show that participants were generally able to effectively filter out repetitive, irrelevant sounds.

While sensory gating was present in both sound conditions, we found that gating strength was affected by task demand when speech was absent, but found no effect when speech was present. In the speech-absent condition, the strength of gating was stronger during the more difficult task, suggesting enhanced suppression of irrelevant auditory stimuli under high task demands. Our findings align with research showing enhanced filtering of irrelevant stimuli at an early stage of processing when task demand was high (Bidet-Caulet, Bottemanne, Fonteneau, Giard, & Bertrand, 2015; Miller, Rietschel, McDonald, & Hatfield, 2011; Sörqvist et al., 2016, 2012). The increased demands of the more difficult task appear to have strengthened sensory filtering, thereby protecting cognitive functioning from the irrelevant stimuli (Lijffijt et al., 2009). Importantly, the paired-click paradigm does not rely on attention markers, such as the P3, which are commonly used in beyond-the-lab studies (Grasso-Cladera, Bremer, Ladouce, & Parada, 2024). Thus, exploring this mechanism across different tasks provides a promising opportunity to investigate how irrelevant sounds are processed without interfering with concurrent tasks (Hölle & Bleichner, 2023b).

When speech was present, we found no effect of task demand on gating strength. One possible explanation is that the speech may have masked the click sounds through energetic masking, where the loudness of the speech obscured the clicks (Brungart, Simpson, Ericson, & Scott, 2001; Hölle & Bleichner, 2023b; Shinn-Cunningham, 2008). This masking likely made the clicks less perceptible, reducing their processing and leading to a lower signal-to-noise ratio (SNR) for the ERPs. This is evident in the descriptively lower ERP amplitudes for each click in the speech-present condition compared to the speech-absent condition. Consequently, the reduced SNR may have decreased the likelihood of finding an effect, as reflected in the weaker gating effect observed in the speech-present compared to speech-absent condition. Alternatively, the speech may have engaged more perceptual resources than the clicks. Stimuli, that are perceived as separate auditory objects, characterized by variations in spatial location or specto-temporal content (Griffiths & Warren, 2004), are also processes separately (Hausfeld, Riecke, Valente, & Formisano, 2018). The separation may lead to a preference to process speech, thereby allocating less resources to the processing of the clicks (Shinn-Cunningham, 2008). This in turn limited the gating effect. Future research could explore these explanations by investigating the influence of irrelevant speech on the processing of other irrelevant sounds.

Utilizing the paired-click paradigm, we investigated how the processing of auditory stimuli was affected by task demand. However, reliance on artificially induced stimuli remains a limitation when transitioning to real-world research (Grasso-Cladera et al., 2024; Matusz et al., 2019). Therefore, we also investigated the processing of the more realistic parts of the soundscape, incorporating the OR playback and speech. Neural responses to irrelevant speech were reduced during high task demand, suggesting that increased workload leave fewer resources available to process speech. Our findings align with previous studies that demonstrated diminished processing of irrelevant stimuli when task demand increased (SanMiguel et al., 2008; Sörqvist et al., 2016, 2012). These studies were conceptually similar to our, as they investigated irrelevant stimuli with a different modality than the task-stimuli. However, these studies used simple tasks and discrete stimuli, such as single tones. Our study extended this work by demonstrating similar effects with continuous and naturalistic stimuli, namely spoken stories. This provides evidence that such mechanisms persist in realistic scenarios. Furthermore, the prolonged allocation of limited cognitive resources to a demanding task while simultaneously suppressing distracting stimuli can be exhausting (Esterman & Rothlein, 2019). This effort is likely to increase when the target and distractor stimuli share the same sensory modality, as competition for cognitive resources increases (Wickens, 2008). For instance, task-relevant speech becomes more difficult to comprehend when presented alongside an OR playback (Way et al., 2013). Investigating how relevant and irrelevant speech compete for cognitive resources while performing another non-auditory task represents a critical step towards understanding distraction in realistic OR scenarios. Previous research has shown larger neural responses to relevant compared to irrelevant speech during a walking task (Straetmans et al., 2024, 2021). Extending such investigations to surgical tasks of varying difficulty would provide insights into the interplay between dual-tasking (i.e., performing a task while processing speech) and speech distraction in high-demand environments like the OR.

The speech suppression effect emerged at an early stage of processing, as indicated by the time-lag of 0 to 300 ms used in the analysis of continuous stimuli. This is similar to the speech literature that usually employ dual- or multi-talker paradigms, showing that ignored speech processing occurs early (Hausfeld, Riecke, Valente, & Formisano, 2018; Hausfeld et al., 2021). This also aligns with the impact of task difficulty on the early ERP components of the paired-click paradigm and suggests that irrelevant speech filtering starts at an early processing stage.

Our exploratory analyses further revealed an inverse relationship between neural responses to speech and self-reported workload, even when accounting for task difficulty. In other words, the strength of the neural response to speech reflected the amount of workload participants experienced. While neurophysiological measures are increasingly used to assess workload, they often rely on ERPs elicited by artificial and repetitive sounds, which are rarely encountered in real-world environments (Wascher et al., 2021). We demonstrated that neural responses to naturalistic stimuli, such as speech, can also serve as markers of perceived workload. This finding could encourage future studies investigating complex work environments to use natural stimuli.

For the OR playback, we found above chance correlations for most participants, showing that stimuli such as an OR playback can be reconstructed from neural data. Contrary to our hypothesis, we found no task modulation for responses to the irrelevant OR playback. While we found a modulated ERP to transient, irrelevant sounds, this did not generalize to all irrelevant sounds. We cannot rule out the possibility that certain sounds within the playback received modulations that were not captured in the computation of a general response to the OR playback. One explanation for this is a discrepancy between our definition of the stimulus categories in our analyses and the actual perception of the OR playback. We defined the clicks, irrelevant speech, and OR playback as separate auditory objects (Griffiths & Warren, 2004). This definition may be applicable to both clicks and speech, as each stimulus exhibited comparable sound characteristics, including amplitude and frequency, over time (Shinn-Cunningham, 2008). However, the OR playback may have been perceived as containing many different auditory objects where each object may have resulted in a different response and task modulation (Huang & Elhilali, 2020). By conceptualizing the entire OR playback as a single auditory object, we may have overlooked sounds within the OR playback that exhibited comparable top-down modulations as observed with the clicks and speech. An alternative explanation is that our analyses may have been biased to enhance speech responses. To have comparable results between the OR playback and speech, we employed the envelope as a feature for both stimuli. This feature is commonly used to maximize speech responses (Crosse et al., 2016). While the envelope is also suitable for describing continuous stimuli beyond speech (e.g., Di Liberto, Pelofi, Shamma, & de Cheveigné, 2020; Hausfeld, Riecke, Valente, & Formisano, 2018; Rosenkranz et al., 2023, 2024), alternative representations, such as the mel-spectogram, explain more variance of the neural data (Di Liberto et al., 2015; Haupt, Rosenkranz, & Bleichner, 2025). Such representations may therefore be more sensitive to detect attention modulations. Either way, when dealing with soundscapes that contain multiple overlapping sounds, a more detailed description, for example by classifying sounds according to their content or acoustic features, may further improve the analyses of responses to the soundscape.

Whilst the present study systematically examined the processing of irrelevant speech and its effect on task performance, certain limitations should be noted. The short duration of the tasks, the performance metrics used, and the relatively high variability of performance in the sample of inexperienced participants may have limited the ability to detect subtle performance impacts (Mentis et al., 2016). Furthermore, the ability to react to distractors is likely to change with experience (Hsu et al., 2008). As the present study focused on participants with no prior experience of the OR, the results may reflect more general demand modulations on auditory processing that are susceptible to change with experience. Hence, future studies utilizing longer tasks and experienced personnel could offer further insights into the impact of speech distraction on surgical personnel.

In the strive to understand auditory distraction in a real-world setting, we combined self-reports, performance, and neurophysiological measures. We investigated how a complex soundscape, consisting of an OR playback, paired clicks, and irrelevant speech is perceived and processed during the performance of surgical tasks that varied in difficulty. Our study demonstrates that while irrelevant speech may not immediately impact performance, it increases perceived workload and distraction during the performance of difficult tasks. This finding could be generalized to other high-stakes settings where a control of the auditory environment is necessary to support the well-being of personnel. Furthermore, our study adds to the growing body of literature that utilizes EEG beyond the lab and at workplaces (e.g., Dehais, Karwowski, & Ayaz, 2020; Grasso-Cladera et al., 2024; Wascher et al., 2021). Using mobile EEG we could investigate the underlying neurophysiological mechanisms of processing the irrelevant aspects of the soundscape, thereby avoiding a direct interference with the surgical task. This showed that irrelevant speech responses were reduced at an early stage of stimulus processing when performing a difficult task. Furthermore, the response was inversely related to workload. This highlights the potential of using mobile EEG to investigate workload in complex, real-world settings with naturalistic stimuli, paving the way for more effective strategies to monitor and mitigate auditory distraction in high-stakes environments.

4.6 Acknowledgement

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) ID 411333557 & 490839860, and by the Forschungspool Funding of the Oldenburg School of Medicine and Health Science. We thank Agnes Hanssen for her help in data acquisition.

4.7 Additional information

The author(s) declare no competing interests. During the preparation of this work the authors used ChatGPT (v. 4) and DeepL Write to improve grammar and wording. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

4.8 Supplementary Material

4.8.1 SURG-TLX items

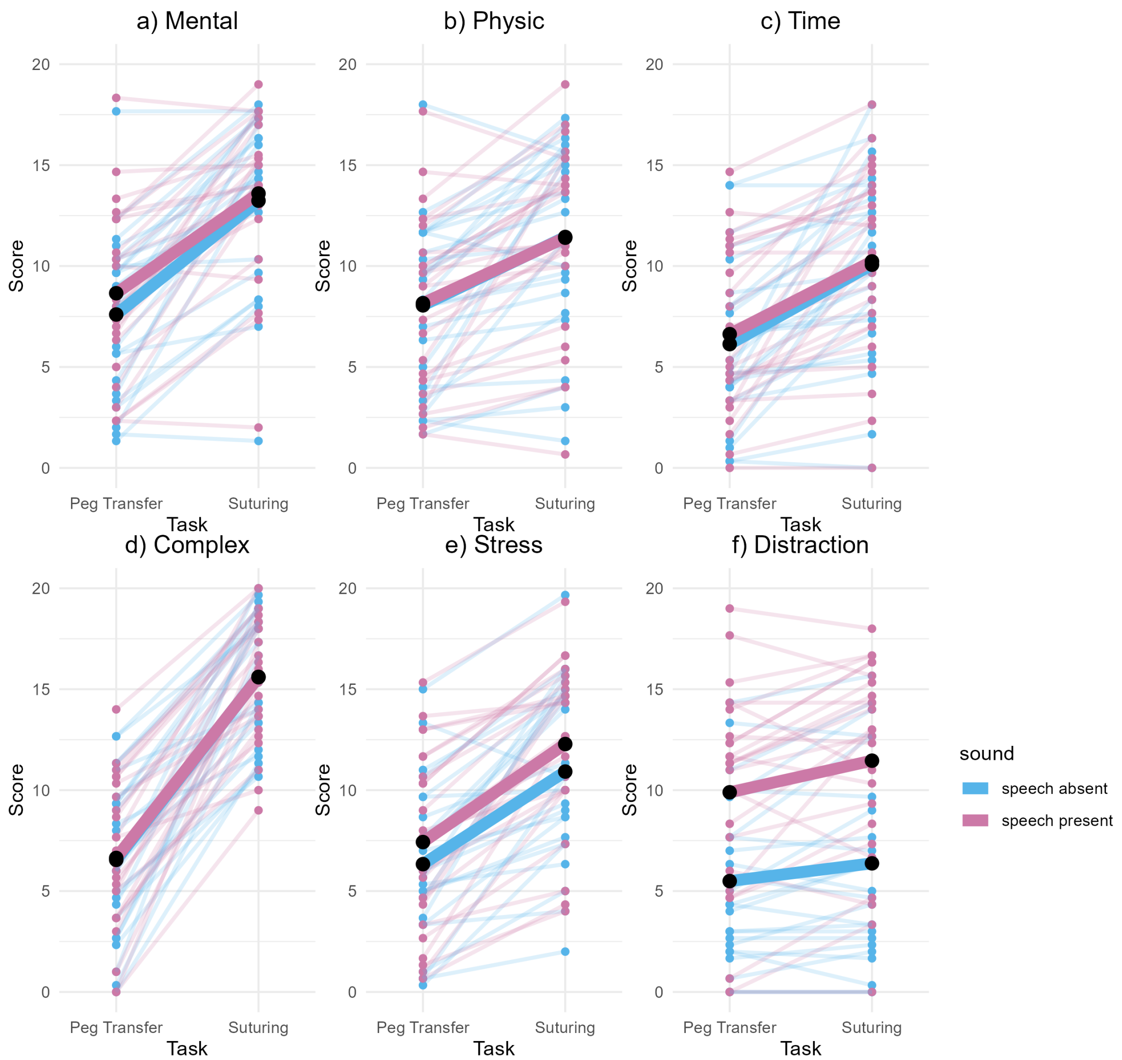


Figure S 4.1: Score for each item of the SURG-TLX for each task and sound condition. The tasks were selected to represent two difficulty levels, with the peg transfer task representing the easy task and suturing representing the difficult task. Score for each item of the SURG-TLX for each task and sound condition. a)-d) showed a significant effect of task, nut not effect of sound condition or an interaction effect. e)-f) showed a significant effect of task and sound condition, but no interaction effect. The thin lines show the participants' data for each sound condition, the thick line the average across participants.

4.8.2 Model comparisons

Outcome	Model Comparison	AIC	Chisq	p-value	Fixed Effects of final model (Estimate $\pm$ SE)
Model Descriptions:					
M0: $\hat{y} \sim 1 + (1\|participant)$					
M1: $\hat{y} \sim task + (1\|participant)$					
M2: $\hat{y} \sim task + sound(1\|participant)$					
M3: $\hat{y} \sim task * sound + (1\|participant)$					
M4: $\hat{y} \sim task + (task\|participant)$					
M5: $\hat{y} \sim task + sound + (task\|participant)$					
M6: $\hat{y} \sim task + sound + (task + sound\|participant)$					
Total	M1 vs M0	473.01	85.86	$< 2.2e-16$ ***	Intercept: 6.69 $\pm$ 0.70 ***
	M2 vs M1	462.99	12.01	0.0005 ***	Task (SU): 4.58 $\pm$ 0.50 ***
	M3 vs M2	464.99	0.0001	0.9934	Sound (Sp): 1.20 $\pm$ 0.27 ***
	M5 vs M2	442.31	24.68	4.37e-06 ***	
	M6 vs M5	438.05	10.26	0.016 *	
Mental	M1 vs M0	515.32	75.04	$< 2.2e-16$ ***	Intercept: 8.13 $\pm$ 0.77 ***
	M2 vs M1	514.92	2.39	0.1219	Task (SU): 5.30 $\pm$ 0.66 ***
	M4 vs M1	504.79	14.52	0.0007 ***	
Physical	M1 vs M0	503.28	47.41	5.75e-12 ***	Intercept: 8.11 $\pm$ 0.82 ***
	M2 vs M1	505.26	0.02	0.8965	Task (SU): 3.33 $\pm$ 0.64 ***
	M4 vs M1	471.30	35.98	1.54e-08 ***	
Time	M1 vs M0	508.78	52.86	3.58e-13 ***	Intercept: 6.38 $\pm$ 0.77 ***
	M2 vs M1	510.09	0.69	0.4054	Task (SU): 3.80 $\pm$ 0.68 ***
	M4 vs M1	478.05	34.73	2.87e-08 ***	
Complexity	M1 vs M0	495.29	143.07	$< 2.2e-16$ ***	Intercept: 6.59 $\pm$ 0.75 ***
	M2 vs M1	497.27	0.02	0.9025	Task (SU): 9.03 $\pm$ 0.74 ***
	M4 vs M1	419.71	79.58	$< 2.2e-16$ ***	
Stress	M1 vs M0	517.15	65.43	6.04e-16 ***	Intercept: 6.25 $\pm$ 0.79 ***
	M2 vs M1	511.49	7.66	0.0056 **	Task (SU): 4.75 $\pm$ 0.62 ***
	M3 vs M2	513.35	0.131	0.716	Sound (Sp): 1.25 $\pm$ 0.43 **
	M5 vs M2	501.91	13.57	0.0011 **	
	M6 vs M5	493.67	14.25	0.0026 **	

Table S 4.1: **Self-reports (except distraction):** Model comparisons for the SURG-TLX total and individual scores. The distraction item can be found in the next table, as the model computation differed for this item from that of the other items. The fixed effects are estimated from the final model, marked in bold. Model descriptions provided at the top of the table. Note that M4 was only tested, if M2 was not significant. SU: Suturing, Sp: Speech present. Significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$.

Outcome	Model Comparison	AIC	Chisq	p-value	Fixed Effects of final model (Estimate $\pm$ SE)
Model Descriptions:					
M0: $\hat{y} \sim 1 + (1 \vert participant)$					
M1: $\hat{y} \sim sound + (1 \vert participant)$					
M2: $\hat{y} \sim sound + task + (1 \vert participant)$					
M3: $\hat{y} \sim sound * task + (1 \vert participant)$					
M4: $\hat{y} \sim sound + task + (sound \vert participant)$					
M5: $\hat{y} \sim sound + task + (sound + task \vert participant)$					
Distraction	M1 vs M0	534.64	60.30	8.15e-15 ***	Intercept: 5.30 $\pm$ 0.92 ***
	M2 vs M1	529.86	6.78	0.0092 **	Sound (Sp): 4.79 $\pm$ 0.48 ***
	M3 vs M2	531.20	0.66	0.4151	Task (SU): 1.27 $\pm$ 0.33 ***
	M4 vs M2	514.49	19.37	6.24e-05 ***	
	M5 vs M4	518.20	2.30	0.513	

Table S 4.2: **Self-report (distraction item):** Model comparisons for the SURG-TLX item distraction. Model descriptions provided at the top of the table. The fixed effects are estimated from the final model, marked in bold. SU: Suturing, Sp: Speech present. Significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$.

Outcome	Model Comparison	AIC	Chisq	p-value	Fixed Effects of final model (Estimate $\pm$ SE)
Model Descriptions:					
M0: $\hat{y} \sim 1 + (1 \vert participant)$					
M1: $\hat{y} \sim sound + (1 \vert participant)$					
PT-Transfers	M1 vs **M0**	281.80	0.0207	0.886	Intercept: 23.013 $\pm$ 1.026 ***
PT-Drops	M1 vs **M0**	248.69	0.1271	0.721	Intercept: 7.1 $\pm$ 0.6 ***
SU-Duration	M1 vs **M0**	-166.98	0.2195	0.64	Intercept: 2.35 $\pm$ 0.26 ***
SU-Damage	M1 vs **M0**	456.49	2.2687	0.132	Intercept: 52.673 $\pm$ 5.016 ***

Table S 4.3: **Surgical task performance:** Model comparisons for the surgical task performance. Model descriptions provided at the top of the table. The fixed effects are estimated from the final model, marked in bold. PT: Peg transfer. SU: Suturing. Significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$.

Outcome	Model Comparison	AIC	Chisq	p-value	Fixed Effects of final model (Estimate $\pm$ SE)
Model Descriptions:					
M0: $\hat{y} \sim 1 + (1 \vert participant)$					
M1: $\hat{y} \sim position + (1 \vert participant)$					
M2: $\hat{y} \sim position + (click \vert participant)$					
amp (Sa)	M1 vs M0	248.89	29.066	6.995e-08 ***	Intercept: 2.97 $\pm$ 0.346 ***
	M2 vs M1	241.87	11.017	0.004 **	position (2nd): -1.16 $\pm$ 0.23 ***
amp (Sp)	M1 vs M0	248.35	7.8145	0.005 **	Intercept: 1.74 $\pm$ 0.227 ***
	M2 vs **M1**	249.68	2.665	0.263	position (2nd): -0.598 $\pm$ 0.21 **

Table S 4.4: **ERP: Presence of gating**: Model comparisons to check whether a gating effect is present. The N1-P2 peak-to-peak amplitude was estimated seperately for the 'speech present' and 'speech absent' condition. Model descriptions provided at the top of the table. The fixed effects are estimated from the final model, marked in bold. Sa: Speech absent, Sp: Speech present. Significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$.

Outcome	Model Comparison	AIC	Chisq	p-value	Fixed Effects of final model (Estimate $\pm$ SE)
Model Descriptions: M0: $\hat{y} \sim 1 + (1\|participant)$ M1: $\hat{y} \sim task + (1\|participant)$ M2: $\hat{y} \sim task + (task\|participant)$					
gating (Sa)	**M1** vs M0 M2 vs M1 (did not converge)	130.24	6.752	0.0093 **	Intercept: 0.764 $\pm$ 0.27 ** Task (SU): 0.79 $\pm$ 0.29 *
gating (Sp)	M1 vs **M0**	140.40	2.48	0.115	Intercept: 0.6 $\pm$ 0.2537 *

Table S 4.5: **ERP: Gating difference between tasks:** Model comparisons to investigate whether the strength of gating differed between tasks. Model descriptions provided at the top of the table. The fixed effects are estimated from the final model, marked in bold. PT: Peg transfer, SU: Suturing, Sa: Speech absent, Sp: Speech present. Significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$.

Outcome	Model Comparison	AIC	Chisq	p-value	Fixed Effects of final model (Estimate $\pm$ SE)
Model Descriptions: M0: $\hat{y} \sim 1 + (1\|participant)$ M1: $\hat{y} \sim task + (1\|participant)$ M2: $\hat{y} \sim task + (task\|participant)$					
playback response (Sa)	M1 vs **M0**	-183.96	2.41	0.121	Intercept: 0.062 $\pm$ 0.006 ***
playback response (Sp)	M1 vs **M0**	-189.77	0.41	0.524	Intercept: 0.045 $\pm$ 0.004 ***
speech response (Sp)	**M1** vs M0 M2 vs M1 (did not converge)	-181.39	11.72	0.001 ***	Intercept: 0.11 $\pm$ 0.01 *** Task (SU): -0.02 $\pm$ 0.004 ***

Table S 4.6: **TRF: Difference between tasks:** Model comparisons to investigate whether the tasks influenced the processing of the different continuous stimuli. Model descriptions provided at the top of the table. The fixed effects are estimated from the final model, marked in bold. SU: Suturing, Sa: Speech absent, Sp: Speech present. Significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$.

Outcome	rc	Model Comparison	AIC	Chisq	p-value	Fixed Effects of final model (Estimate ± SE)
Model Descriptions: M0: $\hat{y} \sim task + (1\|participant)$ M1: $\hat{y} \sim task + rc + (1\|participant)$ M2: $\hat{y} \sim task * rc + (1\|participant)$						
TLX total	Playback response (Sa)	M1 vs **M0** M2 vs M1	212.49 213.96	0.803 1.337	0.370 0.513	Intercept: 7.02 ± 0.81 *** Task (SU): 4.8 ± 0.68 ***
TLX total	Playback response (Sp)	M1 vs **M0** M2 vs M1	210.20 212.20	0.738 0.742	0.390 0.690	Intercept: 8.35 ± 0.83 *** Task (SU): 4.63 ± 0.61 ***
TLX total	speech response (Sp)	M1 vs M0 M2 vs **M1**	206.25 208.21	4.687 0.041	0.030 * 0.840	Intercept: 8.67 ± 0.78 *** Task (SU): 3.98 ± 0.63 *** rc: -1.389 ± 0.637 *

Table S 4.7: **TRF: prediction of workload:** Model comparisons to investigate whether the centered correlation (rc) between the actual and reconstructed stimulus envelope predict self-reported workload. Model descriptions provided at the top of the table. The fixed effects are estimated from the final model, marked in bold. SU: Suturing, Sa: Speech absent, Sp: Speech present. Significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$.

General Discussion 5

5.1 Summary

5.1.1 Study I

In Study I, we investigated how a change in attentional focus influences the processing of auditory stimuli within complex environments. To simulate the complexity of an OR soundscape, we integrated relevant and irrelevant sounds and speech within a Tetris task. Participants were instructed to respond to relevant sounds and speech in order to receive points. The task was also selected due to its requirement for the use of both hands to navigate the blocks, providing a further parallel to surgical tasks. The resulting audio-visual-motor task represented a complex task within a complex soundscape.

The experiment helped to advance the field of auditory stimulus processing, extending beyond the confines of highly controlled laboratory experiments. If a clearly audible alarm and a less audible beep were relevant, we identified a P3 response, indicating an attention shift for both the salient and non-salient sound. Furthermore, we established TRFs as a measure of continuous soundscape processing. The TRF time-course revealed significant time-windows, highlighting the feasibility of measuring TRFs in response to a soundscape comprising relevant and irrelevant speech and sound information while a visual-motor task is being performed. Hence, our study extends the application of TRFs in the study of natural sound processing beyond speech, providing a step towards measurements of audio and EEG beyond the laboratory setting.

5.1.2 Study II

In Study II, we examined how a manipulation of demand during the performance of a surgical task affects the processing of the soundscape. To create a more realistic paradigm in comparison to Study I, we employed a laparoscopic simulator. To manipulate demand, participants performed a serial recall task with two levels of difficulty prior to the surgical task. Additionally, participants were presented with an irrelevant soundscape comprising an OR playback and a changing-state sequence consisting of spoken letters. A further objective of the study was to understand the relationship between the informational content of our stimulus and the neural response. To this end, we computed two stimulus features that carry varying amount of information about the content of the OR playback, namely the envelope and the onsets. We investigated whether they differ in their ability to predict the EEG response.

Participants reported higher workload and distraction for the high demand condition compared to low demand condition. However, an increase in demand in the serial recall task did not result in a decline in performance of the surgical task, suggesting that an increase in secondary task demand and perceived distraction may not necessarily be evident in surgical task performance. Furthermore, ERPs in response to the changing-state sequence, as well as TRFs in response to the OR playback, remained unaltered by demand. However, ERP amplitudes exhibited an effect over time: the N1 amplitude became smaller and the P2 amplitude increased. The null finding for demand may be ascribed to our modification of the serial recall task (i.e. long retention interval and large inter stimulus intervals between spoken letters). The effect of time suggests that responses to varying stimuli, as presented in the changing-state sequence, are subject to modulation. Furthermore, we demonstrated that prediction accuracies of the TRF are similar for the playback envelope and the onsets extracted from the playback. As onsets carry very little information about the content of the OR playback, they could be recorded in the OR without the risk of recording sensitive information. This further increases the applicability of EEG to capture responses to a natural soundscape.

5.1.3 Study III

In Study III, we investigated the impact and processing of irrelevant speech and an OR playback and how it interacts with task demand in a simulated OR environment. To enhance the realism of our approach in comparison with Study II, we varied the task demand using an easy and a difficult surgical task. This mirrors a scenario in the OR where the demand is defined by the surgical task. The experiment incorporated a soundscape consisting of an OR playback, soft clicks, and continuous irrelevant speech. Speech was presented from the back of the participant, mimicking the background speech often encountered in an OR.

The participants self-reported an increase in workload and distraction when engaged in a high demand task compared to a low demand task and when speech was present compared to absent. This finding aligns with the frequently reported distracting effect of irrelevant speech in the OR. As observed in Study II, an increase in distraction ratings did not necessarily result in a decline in performance. Furthermore, we showed that high task demand led to a reduction in the processing of the clicks and irrelevant speech, as measured by ERPs and TRFs, respectively. We found no such effect for the processing of the OR playback. Thus, we replicated that irrelevant sounds (i.e. the clicks) are processed less when task-demand is high compared to low and extended this finding to continuous irrelevant speech.

5.2 Measures of auditory processing and distraction

The distracting nature of the OR soundscape has already been described several decades ago (Shapiro & Berland, 1972). The fact that this is a pressing and unresolved issue is evident from the number of articles that have been published on the topic since we started working on it (e.g. Ayas et al., 2022; Bereuter et al., 2024; Han et al., 2022; Pleban et al., 2021). Our studies contribute to the broader question of how distractors in the OR affect the personnel. Given the limitations of self-reports and performance measures, we were interested whether the processing of auditory distractors could be measured continuously and objectively. Therefore, we focused on the potential of EEG to derive markers of auditory processing by computing ERPs in response to transient sounds and TRFs in response to the soundscapes. In all three studies, we combined EEG with self-reports and performance measurements to receive a comprehensive picture how the soundscape is processed and affects the individual.

5.2.1 ERPs

In all our studies we used ERPs to study responses to transient sounds embedded in continuous soundscapes. The use of established EEG parameters facilitates the interpretation of results when transitioning from controlled laboratory findings to everyday life and workplaces (Gramann et al., 2021; Grasso-Cladera et al., 2024). Our ERP results therefore fit into the neuroergonomic and mobile research field, as we demonstrated that EEG responses to auditory stimuli can be computed in complex environments and situations, providing insights into perceptual and cognitive processes (e.g. Debener et al., 2012; Dehais et al., 2021; Hölle et al., 2021; Protzak & Gramann, 2018; Straetmans et al., 2021; Thomaschewski et al., 2021).

The results of Study I demonstrated that both a relevant and salient alarm, as well as a relevant and soft beep, elicited an attentional response as indicated by the P3 component. Thus, our results are consistent with neuroergonomic and mobile EEG studies, showing that the P3 is a robust marker of attentional processing across various contexts (e.g. Dehais et al., 2021; Grasso-Cladera et al., 2024; Protzak & Gramann, 2018; Suárez et al., 2022). Unlike most studies, we presented relevant sounds sporadically and as part of the task, rather than within the oddball paradigm where a constant auditory stream of irrelevant and relevant stimuli is presented. The constant processing of repetitive sounds provides a rather artificial experimental setting (Grasso-Cladera et al., 2024). Thus, our results could encourage using the P3 component in more realistic setups, for example to investigate attentional trade-offs associated with alarms in the OR. Given that alarms are a significant source of distraction in the OR (Gülşen et al., 2021), largely due to the frequent occurrence of false alarms (Edworthy, 2013), understanding their cognitive impact is crucial. Surgeons may not exhibit overt behavioral responses to false alarms because they recognize that these alarms do not indicate an actual issue (Bliss & Dunn, 2000). However, this reduced behavioral response does not necessarily mean that false alarms do not impose a cognitive burden. Using the P3 in the context of alarm fatigue may shed light on its impact on cognitive resources in high-stakes environments such as the OR.

Interestingly, the P3 showed a rather long drift (see Study I supplementary Figure 2.4), which we did not expect. Similarly, Holtze et al. (2021) observed a late P3 latency when participants heard their own name embedded in an irrelevant speech stream, Protzak and Gramann (2018) found a late P3 latency to an auditory task while participants were driving, and Korte, Jaeger, Rosenkranz, and Bleichner (2024) reported a late P3 latency to natural sounds during task performance. These collective findings suggest that as experimental complexity increases, late cognitive responses can be expected, potentially due to delayed processing of stimuli (Polich, 2007). This highlights the importance of employing established measures, such as ERPs, when transitioning from controlled laboratory settings to real-world research. It provides a basis for ensuring neurophysiological plausibility while also revealing how neural processes may differ in complex, naturalistic environments compared to controlled experimental settings (Grasso-Cladera et al., 2024).

In Study II, we investigated how a changing-state sequence was processed under varying demand. While we found no modulation of demand, we revealed a time-on-task effect in an exploratory analysis. Especially for the N1, adaptation effects, i.e. smaller amplitudes over time, have long been noted (May & Tiitinen, 2010; Näätänen & Picton, 1987). Our results hint to an effect of predictive coding (Friston, 2005), where participants became better at anticipating the occurrence of letters as the experiment progressed. This effect likely arose because the temporal structure of the letters remained consistent across trials, even as their content varied. This would indicate that participants adapted to predictable aspects of the soundscape over time. Translating this to the OR, such findings suggest that experience with the OR environment could allow surgical personnel to adapt to and filter predictable components of the auditory environment, such as ECG monitoring devices, potentially reducing cognitive demand during tasks.

In Study III, we studied ERPs within the framework of sensory gating, a neural mechanism to suppress responses to repetitive and irrelevant stimuli. Our findings align with research demonstrating that the sensory gating effect is robust over long durations, a variety of tasks, and even in the presence of background noise (Hölle & Bleichner, 2023b; Major et al., 2020). Furthermore, we found that high task demand can reduce sensory gating strength, consistent with research linking sensory gating to an individual's ability to shield cognitive functioning from interfering stimuli (Lijffijt et al., 2009). The higher task investment may have reduced the processing of irrelevant and potentially distracting stimuli (Molloy et al., 2019; Sörqvist et al., 2016). While some studies have shown that sensory gating is influenced by cognitive processes such as attention and working memory (e.g., Rosburg, Trautner, Elger, & Kurthen, 2009; Sörqvist et al., 2012), it has not been widely explored in the context of varying task demands. Exploring this effect in other task settings would help to determine whether our results represent a robust and generalizable finding. As such, it is an interesting alternative to the classic oddball paradigm. If the research focus involves studying responses to naturalistic soundscapes, investigating sensory gating may interfere less with the actual soundscape than a continuous oddball paradigm (Hölle & Bleichner, 2023b).

In all our studies, we were mainly interested in the analyses of the irrelevant aspects of a soundscape. Therefore, we focused on early ERP peaks because they are reliably evoked in response to relevant and irrelevant sounds (Näätänen & Picton, 1987). Indeed, we found early responses to all transient stimuli that were the subject of investigation in our studies. Across the studies, I observed a pattern regarding acoustic masking, which describes the effect where the perception of a sound is affected by another sound that is simultaneously presented (Kidd, Mason, Best, & Marrone, 2010; Shinn-Cunningham, 2008): In Study I, the non-salient beep showed a different N1 morphology with a descriptively smaller amplitude than the salient alarm and irrelevant sound. This is partly due to differences in sound intensity, which are known to affect N1 amplitudes (Kaya, Huang, & Elhilali, 2020; Näätänen & Picton, 1987). Additionally, the beep was likely masked by other sounds to an extend that participants often failed to notice the beep, as indicated by the lower hit-rates for the beep. Furthermore, in Study II, the OR playback masked the acoustic onset of most spoken letters, in other words, participants noticed the letter only after a few milliseconds. Likewise, when continuous speech is masked it is processed with a short delay (Brodbeck, Jiao, Hong, & Simon, 2020). To resolve the difference between acoustic and perceptual onset, we shifted the letter onsets to receive time-synchronized responses across different letters. In Study III, the responses to the clicks showed a different morphology and a descriptively smaller gating effect when speech was present compared to when it was absent. This may be due to energetic or informational masking from the speech (Kidd et al., 2010; Shinn-Cunningham, 2008). Overall, when recording EEG responses beyond the presentation of simple stimuli, one must be aware that auditory masking will affect the responses of interest. While this may be desired, for example, to investigate speech processing in noise (Brodbeck et al., 2020), it can also introduce confounds in real-world studies. For example, shifted neural responses due to masking may be mistakenly interpreted as experimental effects rather than as a consequence of overlapping sounds in the environment.

To conclude, our studies contribute to the understanding of how auditory processing operates in complex work environments. By using ERPs to investigate responses to transient sounds embedded in continuous soundscapes, we demonstrated effects of attention (Study I), across time (Study II), and of task-demand (Study III) in dynamic and work-like environments. Thereby, our findings support the potential of ERPs in neuroergonomic research for assessing attentional and perceptual processes (Dehais et al., 2021; Thomaschewski et al., 2021; Wascher et al., 2021).

5.2.2 TRFs

While our ERP findings were indicative that we could compute neurophysiologically plausible responses within the soundscapes, we were also interested in how responses to the entire soundscape can be captured and whether these were affected by the state of a person. In our studies, we computed significant TRF peaks (Study I), and significant prediction accuracies using a forward model (Study II) and backward model (Study III). Thus, our results highlight the potential of TRFs to capture responses to soundscapes in which multiple overlapping sounds are embedded in a continuous stream of auditory information. There are three notable differences from the majority of studies using TRFs that highlight our contribution to the study of continuous stimulus processing:

First, the majority of studies have focused on computing only responses to speech. Utilizing forward models, we computed response to the OR playbacks and found an N1-P2-N2 complex that closely resembled speech responses reported in previous studies (Ding & Simon, 2012a; Fiedler, Wöstmann, Herbst, & Obleser, 2019; Jaeger et al., 2020; Kong et al., 2014). Despite the fundamental difference in content between the OR playbacks and speech, the brain's estimated response to these stimuli may share similarities, especially when using the envelope as a stimulus feature. This is likely because the envelope not only follows the slow amplitude modulations in continuous speech, but also emphasizes its onsets (Crosse et al., 2021; Petersen et al., 2017), a characteristic that is also prominent in OR soundscapes, where many sounds with sharp onsets occur (e.g. alarms and tool clattering). Consequently, the estimated responses to both OR playback and speech are primarily responses of the sensory system to stimulus onsets. Furthermore, sound and speech processing share similar anatomical and functional brain regions, although speech has higher processing demands than non-speech sounds (Price, Thierry, & Griffiths, 2005). This may be reflected in our findings from Study III, where the reconstruction of the OR playback showed descriptively lower correlations compared to the reconstruction of speech. This finding may indicate enhanced speech processing when both stimuli are deemed irrelevant. Since irrelevant speech processing is enhanced when presented alongside babble noise (Herrmann, 2024) or music (Zuk, Murphy, Reilly, & Lalor, 2021), it would be interesting to explore whether this effect extends to other background soundscapes, such as the OR soundscape. Thus, our results not only extend the application of TRFs to continuous stimuli beyond speech, but also demonstrate their ability to differentiate how different aspects of continuous soundscapes are processed.

Second, we investigated soundscapes while participants performed a concurrent non-auditory task. Most studies employing TRFs focus on purely auditory tasks, typically comparing attended and unattended speech (e.g. Brodbeck et al., 2020; Fiedler et al., 2019; Mirkovic et al., 2015; O'Sullivan et al., 2015; Petersen et al., 2017) with few exploring speech during concurrent non-auditory tasks. When participants were engaged in simple visual tasks, such as a visual n-back, TRFs to both relevant and irrelevant speech were successfully recorded (Herrmann, 2024; Xie et al., 2023). In contrast, Vanthornhout et al. (2019) presented irrelevant speech while participants played Tetris. They were unable to compute meaningful responses, which they attributed to artifacts, as meaningful TRFs were obtained in another condition where participants merely watched a movie. Our studies demonstrate that TRFs to irrelevant soundscapes and speech can be reliably computed even when tasks such as Tetris or surgical tasks are performed while sitting or standing. This highlights the utility of TRFs as a measure of auditory processing for complex soundscapes where tasks are concurrently performed. Furthermore, other studies have shown that speech responses can be captured while participants are walking (Straetmans et al., 2024, 2021). Collectively, these findings show that TRFs can capture neural responses to continuous speech and soundscapes across diverse tasks and mobile settings, validating their use as a method that can be applied outside of the lab.

Third, we focused primarily on the processing of the irrelevant aspects of the soundscape. Previous studies have shown that responses to irrelevant speech are typically reduced when attention is directed towards another speech stream (e.g. Mirkovic et al., 2015; O'Sullivan et al., 2015) or a competing task (Vanthornhout et al., 2019). In our studies, we observed that the processing of irrelevant speech varied with task demand, suggesting that speech outside the focus of attention is processed differently depending on task demand. Importantly, while ERP studies with simple stimuli have shown that task demand modulates the processing of irrelevant sounds (Brockhoff et al., 2022), our findings demonstrate a similar effect with natural, continuous stimuli, extending this evidence to a more realistic setting. While we found a modulation of speech, we did not observe similar effects for the OR playbacks, indicating that this modulation may not extend to the soundscape as a whole. Previous studies have shown that stimuli outside the focus of attention are processed separately when their spectro-temporal content differs (Hausfeld, Riecke, Valente, & Formisano, 2018). Consequently, it can be hypothesized that modulations in processing may differ between auditory streams. Investigating how different irrelevant and continuous auditory streams are processed would further increase our understanding of distractor processing. For instance, in the OR, it would be valuable to examine the processing of music. Despite the existence of studies that have reported beneficial effects of music in the OR, the evidence regarding its influence on performance is mixed (Conrad et al., 2012; Han et al., 2022; Kounidas, Kastora, & Maini, 2022; Moorthy et al., 2004). Thus, computing TRFs in response to music (Di Liberto, Pelofi, Bianco, et al., 2020; Di Liberto, Pelofi, Shamma, & de Cheveigné, 2020; Hausfeld, Riecke, Valente, & Formisano, 2018) and the OR soundscape may reveal how music changes the processing of the remaining soundscape. Our results highlight the potential of TRFs to objectively investigated such relationships, thereby providing further insights how irrelevant aspects of the soundscape are processes.

As a further note to this section, I would like to highlight a methodological consideration when computing responses to natural soundscapes, which applies to the OR but also to soundscapes in general: the audio material influences the computation of the TRFs. While this may appear trivial, it is important when recording responses in situations with limited access to a clean, i.e. noise-free, sound source. In Study I, we presented participants with an OR playback from YouTube and could extract the envelope directly from the audio file. In Study II and III, however, we had to apply a noise reduction algorithm before meaningful responses to the envelope could be computed. In these studies we used an OR-recording from the University Hospital Oldenburg (Rennies et al., 2023). We assume that the constant humming sound from ventilation, which was captured in the recording, resulted in an envelope with little variance and an inaccurate estimation of the TRF. Interestingly, using an onset detector algorithm already implemented in a smartphone app (Hölle et al., 2022), we captured onsets directly from the noisy soundscape. The onsets were found to be as predictive of the neural response as the noise-free envelope, underscoring the efficacy of such algorithms even in noisy environments. Since the envelope and onset detector both highlight sound onsets, they are valuable features in an environment like the OR, which is characterized by the presence of numerous transient sounds from machinery and tools. However, it is important to note that they may not be as suitable for soundscapes with a limited number of clear sound onsets. Consequently, while TRFs can be a powerful tool to compute responses, it is crucial to be aware of the characteristics of the soundscape under investigation.

To summarize, our findings demonstrate that TRFs can reliably capture neural responses to complex, naturalistic soundscapes during task performance, extending their application beyond speech research. We also showed that task demand affects the processing of irrelevant speech but not necessarily the broader OR soundscape. Additionally, our results highlight methodological considerations, such as the influence of soundscape characteristics on TRF computation. Together, these findings advance our understanding of auditory processing in complex environments and underscore the potential of TRFs for studying cognitive processes in real-world settings like the OR.

5.2.3 Self-reports and performance

While our primary focus was on EEG measures, we also incorporated subjective and performance measures to gain a comprehensive understanding of the effects of the soundscape. In Studies II and III, we found that perceived distraction did not necessarily impair performance in a surgical task. Previous research has also failed to identify whether performance is affected by the soundscape (Mentis et al., 2016). This underscores the complexity of identifying potentially subtle impacts of auditory distraction on surgical performance. Our results suggest that distraction may elicit a negative emotional response without leading to severe, observable errors. This aligns with the notion that disruptions in the OR might not always translate into immediate negative outcomes for the patient but could still impose cognitive and emotional strain to the personnel (Ayas et al., 2022). Over time, such strain could accumulate as the personnel are required to cope with the auditory environment daily. As prolonged exposure to noise contribute to increased stress (Oiamo, Luginaah, & Baxter, 2015), the soundscape can ultimately pose a risk to the personnel's health.

We further showed that self-reported distraction depends on surgical demand, aligning with observations from the OR (e.g. van Harten et al., 2021). Thus, noise-reduction interventions should be targeted to high-demand phases of a procedure. This finding is also relevant for the training of novice surgeons, as the amount of cognitive resources required by a surgical procedure is dependent on the surgeon's experience level (Hsu et al., 2008; Marrelli et al., 2014; Thomaschewski et al., 2021). For instance, experienced personnel may perceive a procedure as straightforward and therefore the environment as less distracting. Conversely, the same procedure may be highly demanding for an inexperienced person, and thus the environment is perceived as highly distracting (Hsu et al., 2008). Consequently, while assessing the expected demand of a procedure, the level of experience of the surgeons should also be taken into account when planning noise reduction interventions.

In Study III we observed a relationship between task demands, self-reported distraction and neural responses to speech. Speech was perceived as particularly distracting which reinforces its role as a major source of distraction in the OR (Ayas et al., 2022; Healey et al., 2007, 2006; Persoon et al., 2011; Sevdalis et al., 2007, 2014; van Harten et al., 2021). Interestingly, while task demand reduced neural responses to speech, it had no such effect on the overall soundscape. Additionally, speech responses were negatively correlated with self-reported workload, suggesting that participants may have suppressed speech processing by focusing on the task. These findings appear to contradict predictions from load theory (see section 1.2.3), which suggests that higher cognitive load due to a demanding task increases distractor processing (Lavie, 2005). However, this theory has been challenged, as cross-modal distractors, such as those in our study, are more likely to be suppressed under high cognitive load (Brockhoff et al., 2022, 2023; SanMiguel et al., 2008; Sörqvist et al., 2016). Our results support this alternative explanation and demonstrate that such effects can be studied in realistic settings.

Overall, our findings highlight the complex relationship between auditory distraction, neural processing, self-reports, and performance in the OR. While self-reported distraction does not always lead to measurable performance impairments, its cognitive and emotional impact should not be overlooked. The observed reduction in speech processing during high-demand tasks suggests a shielding mechanism to maintain cognitive resources for task performance.

5.3 Limitations

While our studies have contributed to the understanding of auditory processing in complex environments such as the OR, there are certain limitations that occur across all studies that I would like to address in more detail.

In all our studies, the majority of participants were unacquainted with the OR environment and soundscape. While Study I and III included students from various disciplines, the medical students in Study II also had limited experience. It can be expected, that experience shapes the perception and processing of an OR soundscape and therefore our results may not be reflect the processing of an experienced surgeon. As our studies provided initial steps to study auditory processing in the OR, the sample selection was advantageous in terms of accessibility. A sample of experienced surgeons was impractical, given the duration of the studies (three to four hours). However, future studies could use more complex and longer procedures than those used in our studies, which already exist for simulators. For example, based on our experience and research from the speech tracking literature, a 15-minute procedure per condition would be sufficient to obtain reliable responses to an OR soundscape and speech (Crosse et al., 2021; Mesik & Wojtczak, 2023). Consequently, with two conditions and the preparation of EEG, an experiment may last no longer than one hour, which should increase the acceptability of more experienced surgeons to participate in a study.

The soundscapes used in our studies were complex and designed to mimic an actual OR, however, the real OR soundscape is even more complex. A real OR soundscape consists of overlapping relevant and irrelevant sound and speech information. This was considered in Study I, where we computed responses to all sounds in the soundscape, i.e. relevant and irrelevant. However, our studies of auditory processing mostly relate to the irrelevant aspects of the soundscape. As many sounds in the OR are indeed irrelevant for the task and could be avoided (Engelmann et al., 2014), our study contributed to the understanding of these sounds. Nevertheless, to fully capture the complexity of the OR environment, relevant sounds must also be considered. Thus, alarm sounds or speech that must be attended to and reacted to would further increase the representation of a complex environment where multi-tasking is often required.

Throughout the studies, I analyzed responses to the OR playback using the same approach as for speech, which maps features of a continuous auditory stream (e.g. the envelope) onto the neural response, or vice versa (Crosse et al., 2016). Speech is likely processed as a single auditory object, meaning that the same speech stream forms a continuous perceptual unit over time (Ding & Simon, 2012a; Griffiths & Warren, 2004). Auditory attention operates on these objects, enabling suppression of unattended stimuli, as demonstrated in attended and unattended speaker scenarios (Ding & Simon, 2012a; Shinn-Cunningham, 2008; Shinn-Cunningham, Mehraei, Bressler, & Masud, 2013). However, it is debatable whether the OR playback is processed as a single continuous object or as a collection of distinct auditory objects, each eliciting different neural responses. For example, salient events within the OR soundscape, such as the drop of a metal tray or the feedback sound of an electric cutter, likely evoke stronger neural responses compared to less salient background sounds (Huang & Elhilali, 2017, 2020). Treating the OR playback as a continuous object provided a practical and efficient method for capturing neural responses to naturalistic soundscapes. However, this approach may miss the unique contributions of individual sounds within the soundscape.

5.4 Future directions

In order to gain further understanding of how the soundscape affects the personnel in the actual OR and OR-like situations, several aspects could be investigated in future studies.

First, to implement EEG in workplace environments requires technological advances to improve acceptance among personnel. Traditional cap-EEG setups can be uncomfortable to wear over extended periods, limiting their practicality in settings like the OR. To address this, reduced EEG setups that maintain high data quality are essential. For example, ear-EEG which involves small electrode layouts positioned around the ears (Bleichner & Debener, 2017; Debener et al., 2015), has been shown to record high-quality EEG data for several hours (Hölle et al., 2021). Ear-EEG is also sensitive to auditory processing and attention (Bleichner, Mirkovic, & Debener, 2016; Holtze et al., 2022; Meiser & Bleichner, 2022; Meiser et al., 2020). This makes ear-EEG a promising alternative to traditional cap-EEG by offering improved comfort without sacrificing sensitivity to auditory processes. Nevertheless, further comparison of ear-EEG with cap-EEG in challenging, artifact-prone settings is required. For example, during my studies using cap-EEG in a surgical simulation, I encountered muscle artifacts at the mastoid electrodes that had to be removed using ICA. Interestingly, these artifacts were task-specific, as they did not appear during the Tetris task. Such artifact removal may become even more complex with ear-EEG due to its proximity to facial and jaw muscles (Mirkovic, Bleichner, De Vos, & Debener, 2016). Overall, small electrode set-ups like ear-EEG provide an important step for mobile and neuroergonomic assessments but currently require even higher expertise than cap-EEG for both data acquisition (Hölle & Bleichner, 2023a) and processing.

Second, to increase our understanding how different aspects of a soundscape are processed, future studies should combine EEG measurement with acoustic scene analysis to identify and categorize distinct sounds within these environments (e.g. Mesaros, Heittola, Eronen, & Virtanen, 2010; Rennies et al., 2023). A reanalysis of the dataset from Study I revealed that the alarm, irrelevant sounds, and beep accounted for only a small proportion of the overall soundscape, but produced high prediction accuracies (Haupt et al., 2025). Identifying such sounds within a soundscape is essential for understanding their contribution to neural responses. For example, salience detection algorithms could help identify sounds that capture attention (Huang & Elhilali, 2017; Straetmans et al., 2021). However, such algorithms may not be suitable to detect low-salience sounds (Kaya & Elhilali, 2017), like the soft beep in Study I. Alternatively, sounds could be categorized by their content, for example, speech detection algorithms could identify phases of communication in recordings (e.g. Mesaros et al., 2010; Rennies et al., 2023), which could then be related to EEG responses. This would allow the investigation of speech in natural environments. Ideally, such algorithms do not record the raw audio, but features of the audio that guarantee data protection, for example by recording speech features that do not reveal the identity of the speaker (Pohlhausen, Nespoli, & Bitzer, 2024). Combining acoustic scene analysis with EEG will allow researchers to capture neural responses to specific sounds in real-world soundscapes and explore how these are modulated by internal (e.g. cognitive) and external (e.g. acoustic) factors.

Third, our focus was to transition EEG research from laboratory-based studies to more realistic environments. However, this transition is not a straightforward process but rather a dynamic back-and-forth along a continuum between controlled laboratory settings and real-world scenarios (Matusz et al., 2019). Thus, several aspects of this thesis could be further investigated in more controlled settings. For instance, the effect of cognitive load on the processing of a changing-state sequence in a serial-recall task has not been systematically tested (Marois & Vachon, 2024). While Study II addressed this question, our approach deviated from the traditional application of this paradigm, limiting our ability to interpret the findings within the duplex-mechanism account (see section 1.2.1). Investigating how cognitive load interacts with this theory would help determine its applicability across different contexts. Similarly, the reduced processing of irrelevant speech under high task demand observed in our studies could be further examined in controlled settings. Xie et al. (2023) found that relevant speech responses were reduced when participants performed a working memory task (n-back) compared to a no-task condition, though no difference emerged between easy and difficult n-back tasks. A similar study could assess how irrelevant speech is processed under varying cognitive load, providing a more controlled replication of our findings. Additionally, the irrelevant speech effect has primarily been studied using ERPs in response to changing-state sequences (Marois & Vachon, 2024). Applying TRFs to investigate responses to continuous irrelevant speech during a serial-recall task could offer further insights into the neurophysiological processes underlying speech distraction. By integrating theories from controlled experiments with findings and methods from naturalistic designs and vice versa, future research can refine our understanding of auditory distraction, cognitive load, and their neural mechanisms.

Lastly, similarly to the EEG response to objectively measure processing of the OR soundscape, there is also a need for an objective, continuous, and reliable measure of surgical task performance. Given that surgical procedures largely determine task demand, incorporating such measures would enhance the study of distraction in the OR. In Studies II and III, we used performance measures derived from the simulator to objectively assess performance of the surgical task. However, as previously discussed, performance parameters may not be sensitive enough to detect subtle behavioral changes caused by environmental demands. Moreover, capturing performance in a real OR setting poses additional challenges: Direct assessment through surgical outcomes is challenging due to the heterogeneity of surgical procedures (Engelmann et al., 2014). Furthermore, performance assessment through expert evaluations, such as video annotations, are time-consuming, resource-intensive, and introduces privacy concerns. Hence, there is a need for unobtrusive methods that circumvent these issues. One potential solution is to capture movement patterns of the hands, as dexterity may decrease under distraction (Lopus, 2012). This could be achieved by using inertial measurement units (IMUs) positioned on the hands. Similarly, EMG could measure differences in muscle activity, which has been shown to vary between noisy and quiet conditions during surgical tasks (Gao et al., 2018). Identifying a performance measure that minimally disturbs personnel while effectively capturing surgical task performance would be a critical step toward understanding how auditory distractions affect surgical personnel.

5.5 Application

Using EEG to continuously monitor responses to natural soundscapes could provide valuable insights into various work environments and work-like situations. Our findings are particularly relevant for settings characterized by frequent, transient sounds, such as the OR, where numerous auditory events occur that can be linked to EEG. Another setting can, for example, involve aviation personnel who are regularly exposed to beeps, alarms, and speech. The soundscape has been reported to negatively affect performance and well-being (Dehais et al., 2014; Peryer, Noyes, Pleydell-Pearce, & Lieven, 2005). Our studies show that EEG can be used to investigate the processing of auditory stimuli in such situations. Thus, in such workplaces, transient sounds can serve as natural events that provide information about cognitive processes (Wascher et al., 2021).

Computing EEG responses to the continuous OR soundscape could also be useful to evaluate the effectiveness of training programs that incorporate realistic auditory environments. Suárez et al. (2022) and Thomaschewski et al. (2021) have shown that laparoscopic training alters brain activity, with EEG indicating reduced demands as proficiency increases. Since our results suggest that task demand influences the processing of specific aspects of the soundscape, such as speech, TRF responses could reveal how experience-driven changes in task demand affect the neural processing of both speech and the OR soundscape. This approach could provide insight into how well training prepares personnel for the auditory and cognitive demands of the OR.

Furthermore, EEG could be a valuable tool for assessing intervention strategies aimed at reducing distraction. Leitsmann et al. (2021) showed that a communication device designed to suppress irrelevant noise while enhancing relevant speech had self-reported benefits but no effect on performance. Our findings are promising to evaluate the cognitive impact of such interventions. For example, examining how irrelevant speech that is suppressed by such a device is processed under varying task demands could help determine whether these interventions effectively reduce cognitive load. Additionally, any noise reduction strategy must ensure that critical auditory signals, such as alarms, remain perceptible. Measuring ERPs to alarms could indicate whether they are processed or whether personnel fail to respond, possibly due perceptual failure (Dehais et al., 2014). This approach could guide improvements in OR sound design, leading to alarms that effectively capture attention while minimizing disruption to surgical personnel (Anderson, Sreetharan, Elizondo López, Schlesinger, & Schutz, 2023).

Finally, this thesis not only advanced the use of EEG in work-like settings but also provides insights into strategies for reducing distraction in the OR. Noise reduction training (e.g., Engelmann et al., 2014) should emphasize the importance of minimizing unnecessary noise, particularly during high demand phases of surgery. Our findings align with observations that higher task demand is associated with increased perceived distraction (van Harten et al., 2021), underscoring the need to limit irrelevant sounds during these periods. In particular, case-irrelevant speech, such as teaching, phone calls or private conversation, should be avoided when task demand is high. However, during phases of low demand, such speech may be less disruptive and could even help alleviate tension and improve team morale (Ayas et al., 2022; Widmer et al., 2018). Managing when and how case-irrelevant communication occurs could support both noise reduction and team dynamics in the OR.

5.6 Conclusion

In this thesis I investigated how the brain processes complex soundscapes in demanding work environments, with a focus on OR soundscapes. Using EEG, we examined how soundscapes, including a variety of overlapping sounds such as alarms and irrelevant speech, are processed and how they relate to perceived workload and performance during complex tasks. Across three studies, we demonstrated that ERPs reliably capture transient auditory responses, while TRFs provide a valuable tool for assessing neural responses to continuous soundscapes. Importantly, we showed that TRFs can capture responses to irrelevant soundscapes beyond speech, even while a task is being performed. Additionally, we demonstrated that neural responses in such complex settings evolve over time and that processing of irrelevant speech is modulated by surgical task demand. Our findings further revealed that self-reported distraction and surgical performance are not necessarily related, reinforcing the need for measures beyond task performance to accurately assess distraction in the OR. By extending EEG research beyond highly controlled laboratory settings, we successfully recorded reliable neural responses to sound, thereby investigating traditional laboratory findings in more naturalistic environments. I encourage future research to continue to bridge the gap between controlled experiments and real-world studies. By advancing along this continuum, we helped to drive methodological developments, identify potential theoretical limitations, and ultimately improved our understanding of the interaction between auditory and cognitive processes in high-stakes environments.

References

Aiken, S. J., & Picton, T. W. (2008). Human cortical responses to the speech envelope. *Ear and Hearing*, *29*(2), 139–157. doi: 10.1097/AUD.0b013e31816453dc

Anderson, C. J., Sreetharan, S., Elizondo López, A. E., Schlesinger, J. J., & Schutz, M. (2023). Improving auditory alarms: Reducing perceived annoyance with musical timbre (a randomized trial). *Perioperative Care and Operating Room Management*, *32*, 100332. doi: 10.1016/j.pcorm.2023.100332

Arora, S., Sevdalis, N., Nestel, D., Woloshynowych, M., Darzi, A., & Kneebone, R. (2010). The impact of stress on surgical performance: A systematic review of the literature. *Surgery*, *147*(3), 318-330.e6. doi: 10.1016/j.surg.2009.10.007

Ayas, S., Donmez, B., Kazlovich, K., Lombardi, S., & Jain, A. (2022). The Prevalence and Potential Effects of Distractions in General Surgery: A Mixed-Methods Study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *66*(1), 2132–2136. doi: 10.1177/1071181322661335

Baltin, C. T., Wilhelm, H., Wittland, M., Hoelscher, A. H., Stippel, D., & Astvatsatourov, A. (2020). Noise patterns in visceral surgical procedures: Analysis of second-by-second dBA data of 599 procedures over the course of one year. *Scientific Reports*, *10*(1), 1–10. doi: 10.1038/s41598-020-59816-4

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). doi: 10.18637/jss.v067.i01

Bereuter, J.-P., Geissler, M. E., Geissler, R., Schmidt, S., Buck, N., Weiß, J., . . . von Bechtolsheim, F. (2024). Impact of Operative Room Noise on Laparoscopic Performance - A Prospective, Randomized Crossover Trial. *Journal of Surgical Research*, *302*, 648–655. doi: 10.1016/j.jss.2024.07.110

Berglund, B., Lindvall, T., Schwela, D. H., & World Health Organization Occupational and Environmental Health Team. (1999). Guidelines for community noise. *World Health Organization*.

Bidet-Caulet, A., Bottemanne, L., Fonteneau, C., Giard, M. H., & Bertrand, O. (2015). Brain Dynamics of Distractibility: Interaction Between Top-Down and Bottom-Up Mechanisms of Auditory Attention. *Brain Topography*, *28*(3), 423–436. doi: 10.1007/s10548-014-0354-x

Bleichner, M. G., & Debener, S. (2017). Concealed, unobtrusive ear-centered EEG acquisition: Ceegrids for transparent EEG. *Frontiers in Human Neuroscience*, *11*(April), 1–14. doi: 10.3389/fnhum.2017.00163

Bleichner, M. G., Mirkovic, B., & Debener, S. (2016). Identifying auditory attention with ear-EEG: CEEGrid versus high-density cap-EEG comparison. *Journal of Neural Engineering*, *13*(6). doi: 10.1088/1741-2560/13/6/066004

Bliss, J. P., & Dunn, M. C. (2000). Behavioural implications of alarm mistrust as a function of task workload. *Ergonomics*, *43*(9), 1283–1300. doi: 10.1080/001401300421743

Blum, S., Hölle, D., Bleichner, M. G., & Debener, S. (2021). Pocketable Labs for Everyone: Synchronized Multi-Sensor Data Streaming and Recording on Smartphones with the Lab Streaming Layer. *Sensors*, *21*(23), 8135. doi: 10.3390/s21238135

Bohrer, T., Koller, M., Schlitt, H. J., Bauer, H., & on behalf of the German Society of Surgery. (2011). Workload and quality of life of surgeons. Results and implications of a large-scale survey by the German Society of Surgery. *Langenbeck's Archives of Surgery*, *396*(5), 669–676. doi: 10.1007/s00423-011-0791-x

Broadbent, D. E. (1958). *Perception and communication*. Oxford: Oxford University Press.

Brockhoff, L., Schindler, S., Bruchmann, M., & Straube, T. (2022). Effects of perceptual and working memory load on brain responses to task-irrelevant stimuli: Review and implications for future research. *Neuroscience & Biobehavioral Reviews*, *135*, 104580. doi: 10.1016/j.neubiorev.2022.104580

Brockhoff, L., Vetter, L., Bruchmann, M., Schindler, S., Moeck, R., & Straube, T. (2023). The effects of visual working memory load on detection and neural processing of task-unrelated auditory stimuli. *Scientific Reports*, *13*(1), 4342. doi: 10.1038/s41598-023-31132-7

Brodbeck, C., Jiao, A., Hong, L. E., & Simon, J. Z. (2020). Neural speech restoration at the cocktail party: Auditory cortex recovers masked speech of both attended and ignored speakers. *PLOS Biology*, *18*(10), e3000883. doi: 10.1371/journal.pbio.3000883

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*, *110*(5), 2527–2538. doi: 10.1121/1.1408946

Campbell, T., Winkler, I., Kujala, T., & Näätänen, R. (2003). The N1 hypothesis and irrelevant sound: Evidence from token set size effects. *Cognitive Brain Research*, *18*(1), 39.

Cohen, M. X. (2022). A tutorial on generalized eigendecomposition for denoising, contrast enhancement, and dimension reduction in multichannel electrophysiology. *NeuroImage*, *247*, 118809. doi: 10.1016/j.neuroimage.2021.118809

Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behavior*, *15*(1), 17–31. doi: 10.1016/S0022-5371(76)90003-7

Conrad, C., Konuk, Y., Werner, P., Cao, C. G., Warshaw, A., Rattner, D., ... Gee, D. (2010). The effect of defined auditory conditions versus mental loading on the laparoscopic motor skill performance of experts. *Surgical Endoscopy*, *24*(6), 1347–1352. doi: 10.1007/s00464-009-0772-0

Conrad, C., Konuk, Y., Werner, P. D., Cao, C. G., Warshaw, A. L., Rattner, D. W., ... Gee, D. W. (2012). A quality improvement study on avoidable stressors and countermeasures affecting surgical motor performance and learning. *Annals of Surgery*, *255*(6), 1190–1194. doi: 10.1097/SLA.0b013e318250b332

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215. doi: 10.1038/nrn755

Crosse, M. J., Butler, J. S., & Lalor, E. C. (2015). Congruent Visual Speech Enhances Cortical Entrainment to Continuous Auditory Speech in Noise-Free Conditions. *Journal of Neuroscience*, *35*(42), 14195–14204. doi: 10.1523/JNEUROSCI.1829-15.2015

Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The multivariate temporal response function (mTRF) toolbox: A MATLAB toolbox for relating neural signals to continuous stimuli. *Frontiers in Human Neuroscience*, *10*(NOV2016), 1–14. doi: 10.3389/fnhum.2016.00604

Crosse, M. J., Zuk, N. J., Di Liberto, G. M., Nidiffer, A. R., Molholm, S., & Lalor, E. C. (2021). Linear Modeling of Neurophysiological Responses to Speech and Other Continuous Stimuli: Methodological Considerations for Applied Research. *Frontiers in Neuroscience*, *15*, 705621. doi: 10.3389/fnins.2021.705621

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: Age, sleep and modality. *Clinical Neurophysiology*, *115*(4), 732–744. doi: 10.1016/j.clinph.2003.11.021

Czerwiec, A., Vannier, M., & Courage, O. (2024). Effect of noise on the performance of arthroscopic simulator. *Surgery Open Science*, *20*, 145–150. doi: 10.1016/j.sopen.2024.06.006

Daeglau, M., Otten, J., Grimm, G., Mirkovic, B., Hohmann, V., & Debener, S. (2025). *Neural Speech Tracking in a Virtual Acoustic Environment: Audio-Visual Benefit for Unscripted Continuous Speech* (No. arXiv:2501.08124). arXiv. doi: 10.48550/arXiv.2501.08124

Daeglau, M., Otten, J., Mirkovic, B., Grimm, G., Debener, S., & Hohmann, V. (2023). *Audiovisual recordings of unscripted monologues.* Zenodo. doi: 10.5281/zenodo.8082844

Debener, S., Emkes, R., De Vos, M., & Bleichner, M. (2015). Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear. *Scientific Reports*, *5*(November). doi: 10.1038/srep16743

Debener, S., Minow, F., Emkes, R., Gandras, K., & de Vos, M. (2012). How about taking a low-cost, small, and wireless EEG for a walk? *Psychophysiology*, *49*(11), 1617–1621. doi: 10.1111/j.1469-8986.2012.01471.x

DeCaporale-Ryan, L., Sakran, J. V., Grant, S. B., Alseidi, A., Rosenberg, T., Goldberg, R. F., ... Moalem, J. (2017). The undiagnosed pandemic: Burnout and depression within the surgical community. *Current Problems in Surgery*, *54*(9), 453–502. doi: 10.1067/j.cpsurg.2017.07.001

Dehais, F., Causse, M., Vachon, F., Régis, N., Menant, E., & Tremblay, S. (2014). Failure to detect critical auditory alerts in the cockpit: Evidence for inattentional deafness. *Human Factors*, *56*(4), 631–644. doi: 10.1177/0018720813510735

Dehais, F., Karwowski, W., & Ayaz, H. (2020). Brain at Work and in Everyday Life as the Next Frontier: Grand Field Challenges for Neuroergonomics. *Frontiers in Neuroergonomics*, *1*(October), 1–9. doi: 10.3389/fnrgo.2020.583733

Dehais, F., Somon, B., Mullen, T., & Callan, D. E. (2021). A Neuroergonomics Approach to Measure Pilot's Cognitive Incapacitation in the Real World with EEG. In H. Ayaz & U. Asgher (Eds.), *Advances in Neuroergonomics and Cognitive Engineering* (pp. 111–117). Cham: Springer International Publishing. doi: 10.1007/978-3-030-51041-1_16

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods*, *13*, 9–21. doi: http://dx.doi.org/10.1016/j.jneumeth.2003.10.009

De Vos, M., Gandras, K., & Debener, S. (2014). Towards a truly mobile auditory brain-computer interface: Exploring the P300 to take away. *International Journal of Psychophysiology*, *91*(1), 46–53. doi: 10.1016/j.ijpsycho.2013.08.010

Dias, R. D., Ngo-Howard, M. C., Boskovski, M. T., Zenati, M. A., & Yule, S. J. (2018). Systematic review of measurement tools to assess surgeons' intraoperative cognitive workload. *British Journal of Surgery*, *105*(5), 491–501. doi: 10.1002/bjs.10795

Di Liberto, G. M., O'Sullivan, J. A., & Lalor, E. C. (2015). Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. *Current biology: CB*, *25*(19), 2457–2465. doi: 10.1016/j.cub.2015.08.030

Di Liberto, G. M., Pelofi, C., Bianco, R., Patel, P., Mehta, A. D., Herrero, J. L., ... Mesgarani, N. (2020). Cortical encoding of melodic expectations in human temporal cortex. *eLife*, *9*, e51784. doi: 10.7554/eLife.51784

Di Liberto, G. M., Pelofi, C., Shamma, S., & de Cheveigné, A. (2020). Musical expertise enhances the cortical tracking of the acoustic envelope during naturalistic music listening. *Acoustical Science and Technology*, *41*(1), 361–364. doi: 10.1250/ast.41.361

Ding, N., & Simon, J. Z. (2012a). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(29), 11854–11859. doi: 10.1073/pnas.1205381109

Ding, N., & Simon, J. Z. (2012b). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology*, *107*(1), 78–89. doi: 10.1152/jn.00297.2011

Drennan, D. P., & Lalor, E. C. (2019). Cortical Tracking of Complex Sound Envelopes: Modeling the Changes in Response with Intensity. *eNeuro*, *6*(3). doi: 10.1523/ENEURO.0082-19.2019

Edworthy, J. (2013). Medical audible alarms: A review. *Journal of the American Medical Informatics Association*, *20*(3), 584–589. doi: 10.1136/amiajnl-2012-001061

Ellermeier, W., & Zimmer, K. (2014). The psychoacoustics of the irrelevant sound effect. *Acoustical Science and Technology*, *35*(1), 10–16. doi: 10.1250/ast.35.10

Engelmann, C. R., Neis, J. P., Kirschbaum, C., Grote, G., & Ure, B. M. (2014). A noise-reduction program in a pediatric operation theatre is associated with surgeon's benefits and a reduced rate of complications: A prospective controlled clinical trial. *Annals of Surgery*, *259*(5), 1025–1033. doi: 10.1097/SLA.0000000000000253

Esterman, M., & Rothlein, D. (2019). Models of sustained attention. *Current Opinion in Psychology*, *29*, 174–180. doi: 10.1016/j.copsyc.2019.03.005

Etheridge, J. C., Evans, D., Zhao, L., Ibrahim, N., Wick, E. C., Freischlag, J. A., & Brownstein, M. R. (2023). Trends in Surgeon Burnout in the US and Canada: Systematic Review and Meta-Regression Analysis. *Journal of the American College of Surgeons*, *236*(1), 253. doi: 10.1097/XCS.0000000000000402

Feuerbacher, R. L., Funk, K. H., Spight, D. H., Diggs, B. S., & Hunter, J. G. (2012). Realistic Distractions and Interruptions That Impair Simulated Surgical Performance by Novice Surgeons. *Archives of Surgery*, *147*(11), 1026–1030. doi: 10.1001/archsurg.2012.1480

Fiedler, L., Wöstmann, M., Herbst, S. K., & Obleser, J. (2019). Late cortical tracking of ignored speech facilitates neural selectivity in acoustically challenging conditions. *NeuroImage*, *186*, 33–42. doi: 10.1016/j.neuroimage.2018.10.057

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. doi: 10.1098/rstb.2005.1622

Gao, J., Liu, S., Feng, Q., Zhang, X., Jiang, M., Wang, L., ... Zhang, Q. (2019). Subjective and Objective Quantification of the Effect of Distraction on Physician's Workload and Performance During Simulated Laparoscopic Surgery. *Medical Science Monitor*, *25*, 3127–3132. doi: 10.12659/MSM.914635

Gao, J., Liu, S., Feng, Q., Zhang, X., Zhang, J., Jiang, M., ... Zhang, Q. (2018). Quantitative Evaluations of the Effects of Noise on Mental Workloads Based on Pupil Dilation during Laparoscopic Surgery. *The American Surgeon*, *84*(12), 1951–1956. doi: 10.1177/000313481808401243

Georgiou, K., Larentzakis, A., & Papavassiliou, A. G. (2017). Surgeons' and surgical trainees' acute stress in real operations or simulation: A systematic review. *The Surgeon*, *15*(6), 355–365. doi: 10.1016/j.surge.2017.06.003

Giv, M. D., Sani, K. G., Alizadeh, M., Valinejadi, A., & Majdabadi, H. A. (2017). Evaluation of noise pollution level in the operating rooms of hospitals: A study in Iran. *Interventional Medicine & Applied Science*, *9*(2), 61–66. doi: 10.1556/1646.9.2017.15

Gonthier, C. (2022). An easy way to improve scoring of memory span tasks: The edit distance, beyond "correct recall in the correct serial position". *Behavior Research Methods*. doi: 10.3758/s13428-022-01908-2

Gramann, K., Ferris, D. P., Gwin, J., & Makeig, S. (2014). Imaging natural cognition in action. *International Journal of Psychophysiology*, *91*(1), 22–29. doi: 10.1016/j.ijpsycho.2013.09.003

Gramann, K., McKendrick, R., Baldwin, C., Roy, R. N., Jeunet, C., Mehta, R. K., & Vecchiato, G. (2021). Grand Field Challenges for Cognitive Neuroergonomics in the Coming Decade. *Frontiers in Neuroergonomics*, *2*(March). doi: 10.3389/fnrgo.2021.643969

Grasso-Cladera, A., Bremer, M., Ladouce, S., & Parada, F. (2024). A systematic review of mobile brain/body imaging studies using the P300 event-related potentials to investigate cognition beyond the laboratory. *Cognitive, Affective, & Behavioral Neuroscience*, *24*(4), 631–659. doi: 10.3758/s13415-024-01190-z

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience*, *5*(11), 887–892. doi: 10.1038/nrn1538

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, *48*(12), 1711–1725. doi: 10.1111/j.1469-8986.2011.01273.x

Gülşen, M., Aydıngülü, N., & Arslan, S. (2021). Physiological and psychological effects of ambient noise in operating room on medical staff. *ANZ Journal of Surgery*, *91*(5), 847–853. doi: 10.1111/ans.16582

Haapakangas, A., Hongisto, V., & Liebl, A. (2020). The relation between the intelligibility of irrelevant speech and cognitive performance—A revised model based on laboratory studies. *Indoor Air*, *30*(6), 1130–1146. doi: 10.1111/ina.12726

Hamilton, L. S., & Huth, A. G. (2020). The revolution will not be controlled: Natural stimuli in speech neuroscience. *Language, Cognition and Neuroscience*, *35*(5), 573–582. doi: 10.1080/23273798.2018.1499946

Han, Y., Zheng, B., Zhao, L., Hu, J., Zhang, C., Xiao, R., ... Pu, D. (2022). Impact of background music on the performance of laparoscopy teams. *BMC Medical Education*, *22*(1), 439. doi: 10.1186/s12909-022-03503-7

Hansen, J. C., & Hillyard, S. A. (1980). Endogeneous brain potentials associated with selective auditory attention. *Electroencephalography and Clinical Neurophysiology*, *49*(3), 277–290. doi: 10.1016/0013-4694(80)90222-9

Hari, R., Sams, M., & Järvilehto, T. (1979). Auditory evoked transient and sustained potentials in the human EEG: I. Effects of expectation of stimuli. *Psychiatry Research*, *1*(3), 297–306. doi: 10.1016/0165-1781(79)90011-8

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, *52*, 139–183. doi: 10.1016/S0166-4115(08)62386-9

Hasfeldt, D., Laerkner, E., & Birkelund, R. (2010). Noise in the Operating Room-What Do We Know? A Review of the Literature. *Journal of Perianesthesia Nursing*, *25*(6), 380–386. doi: 10.1016/j.jopan.2010.10.001

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J. D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, *87*, 96–110. doi: 10.1016/j.neuroimage.2013.10.067

Haupt, T., Rosenkranz, M., & Bleichner, M. (2025). Exploring Relevant Features for EEG-Based Investigation of Sound Perception in Naturalistic Soundscapes. *eNeuro*. doi: 10.1523/ENEURO.0287-24.2024

Haupt, T., Rosenkranz, M., & Bleichner, M. G. (2024). *Exploring relevant Features for EEG-Based Investigation of Sound Perception in Naturalistic Soundscapes* [Preprint]. PsyArXiv. Retrieved from https://doi.org/10.31234/osf.io/nuy7e doi: https://doi.org/10.31234/osf.io/nuy7e

Hausfeld, L., Riecke, L., & Formisano, E. (2018). Acoustic and higher-level representations of naturalistic auditory scenes in human auditory and frontal cortex. *NeuroImage*, *173*, 472–483. doi: 10.1016/j.neuroimage.2018.02.065

Hausfeld, L., Riecke, L., Valente, G., & Formisano, E. (2018). Cortical tracking of multiple streams outside the focus of attention in naturalistic auditory scenes. *NeuroImage*, *181*(March), 617–626. doi: 10.1016/j.neuroimage.2018.07.052

Hausfeld, L., Shiell, M., Formisano, E., & Riecke, L. (2021). Cortical processing of distracting speech in noisy auditory scenes depends on perceptual demand. *NeuroImage*, *228*(May 2020), 117670. doi: 10.1016/j.neuroimage.2020.117670

Healey, A. N., Primus, C. P., & Koutantji, M. (2007). Quantifying distraction and interruption in urological surgery. *Quality and Safety in Health Care*, *16*(2), 135–139. doi: 10.1136/qshc.2006.019711

Healey, A. N., Sevdalis, N., & Vincent, C. A. (2006). Measuring intra-operative interference from distraction and interruption observed in the operating theatre. *Ergonomics*, *49*(5-6), 589–604. doi: 10.1080/00140130600568899

Hedman, L., Klingberg, T., Enochsson, L., Kjellin, A., & Felländer-Tsai, L. (2007). Visual working memory influences the performance in virtual image–guided surgical intervention. *Surgical Endoscopy*, *21*(11), 2044–2050. doi: 10.1007/s00464-007-9287-8

Herrmann, B. (2024). Minimal background noise enhances neural speech tracking: Evidence of stochastic resonance. *eLife*, *13*. doi: 10.7554/eLife.100830.1

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. (1973). Electrical Signs of Selective Attention in the Human Brain Abstract. *Science*, *182*(October), 177–179.

Hodge, B., & Thompson, J. F. (1990). Noise pollution in the operating theatre. *The Lancet*, *335*(8694), 891–894. doi: 10.1016/0140-6736(90)90486-O

Hölle, D., & Bleichner, M. G. (2023a). Recording Brain Activity with Ear-Electroencephalography. *Journal of Visualized Experiments*(193), 64897. doi: 10.3791/64897

Hölle, D., & Bleichner, M. G. (2023b). Smartphone-based ear-electroencephalography to study sound processing in everyday life. *European Journal of Neuroscience*, *n/a*(n/a), 1–15. doi:

10.1111/ejn.16124
Hölle, D., Blum, S., Kissner, S., Debener, S., & Bleichner, M. G. (2022). Real-time audio processing of real-life soundscapes for EEG analysis: ERPs based on natural sound onsets. *Frontiers in Neuroergonomics*, *3*. doi: 10.3389/fnrgo.2022.793061
Hölle, D., Meekes, J., & Bleichner, M. G. (2021). Mobile ear-EEG to study auditory attention in everyday life: Auditory attention in everyday life. *Behavior Research Methods*, *53*(5), 2025–2036. doi: 10.3758/s13428-021-01538-0
Holtze, B., Jaeger, M., Debener, S., Adiloğlu, K., & Mirkovic, B. (2021). Are They Calling My Name? Attention Capture Is Reflected in the Neural Tracking of Attended and Ignored Speech. *Frontiers in Neuroscience*, *15*. doi: 10.3389/fnins.2021.643705
Holtze, B., Rosenkranz, M., Jaeger, M., Debener, S., & Mirkovic, B. (2022). Ear-EEG Measures of Auditory Attention to Continuous Speech. *Frontiers in Neuroscience*, *16*.
Horton, C., D'Zmura, M., & Srinivasan, R. (2013). Suppression of competing speech through entrainment of cortical oscillations. *Journal of Neurophysiology*, *109*(12), 3082–3093. doi: 10.1152/jn.01026.2012
Hsu, K. E., Man, F. Y., Gizicki, R. A., Feldman, L. S., & Fried, G. M. (2008). Experienced surgeons can do more than one thing at a time: Effect of distraction on performance of a simple laparoscopic and cognitive task by experienced and novice surgeons. *Surgical Endoscopy and Other Interventional Techniques*, *22*(1), 196–201. doi: 10.1007/s00464-007-9452-0
Hu, J. S. L., Lu, J., Tan, W. B., & Lomanto, D. (2016). Training improves laparoscopic tasks performance and decreases operator workload. *Surgical Endoscopy*, *30*(5), 1742–1746. doi: 10.1007/s00464-015-4410-8
Huang, N., & Elhilali, M. (2017). Auditory salience using natural soundscapes. *The Journal of the Acoustical Society of America*, *141*(3), 2163–2176. doi: 10.1121/1.4979055
Huang, N., & Elhilali, M. (2020). Push-pull competition between bottom-up and top-down auditory attention to natural soundscapes. *eLife*, *9*, 1–22. doi: 10.7554/eLife.52984
Hughes, R. W. (2014). Auditory distraction: A duplex-mechanism account. *PsyCh Journal*, *3*(1), 30–41. doi: 10.1002/pchj.44
Jacobsen, N. S. J., Blum, S., Witt, K., & Debener, S. (2021). A walk in the park? Characterizing gait-related artifacts in mobile EEG recordings. *European Journal of Neuroscience*, *54*(12), 8421–8440. doi: 10.1111/ejn.14965
Jaeger, M., Mirkovic, B., Bleichner, M. G., & Debener, S. (2020). Decoding the Attended Speaker From EEG Using Adaptive Evaluation Intervals Captures Fluctuations in Attentional Listening. *Frontiers in Neuroscience*, *14*, 603. doi: 10.3389/fnins.2020.00603
Jeyaraman, M., Jeyaraman, N., Yadav, S., Nallakumarasamy, A., Iyengar, K. P., Jain, V., ... Jain, V. (2024). Impact of Excessive Noise Generation in Orthopaedic Operating Theatres: A Comprehensive Review. *Cureus*, *16*. doi: 10.7759/cureus.54469
Jones, D., Madden, C., & Miles, C. (1992). Privileged Access by Irrelevant Speech to Short-term Memory: The Role of Changing State. *The Quarterly Journal of Experimental Psychology Section A*, *44*(4), 645–669. doi: 10.1080/14640749208401304
Jones, D., & Morris, N. (1992). Irrelevant speech and serial recall: Implications for theories of attention and working memory. *Scandinavian Journal of Psychology*, *33*(3), 212–229. doi: 10.1111/j.1467-9450.1992.tb00911.x
Jung, J. J., Jüni, P., Lebovic, G., & Grantcharov, T. (2020). First-year Analysis of the Operating Room Black Box Study. *Annals of Surgery*, *271*(1), 122–127. doi: 10.1097/SLA.0000000000002863
Kalarus, M. (2021). *Tetris3d.* https://www.mathworks.com/matlabcentral/fileexchange/92243-tetris3d. ([Online; accessed September, 2021])
Kattner, F., Föcker, J., Moshona, C. C., & Marsh, J. E. (2024). When softer sounds are more distracting: Task-irrelevant whispered speech causes disruption of serial recall. *The Journal of the Acoustical Society of America*, *156*(5), 3632–3648. doi: 10.1121/10.0034454
Kaya, E. M., & Elhilali, M. (2017). Modelling auditory attention. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1714). doi: 10.1098/rstb.2016.0101

Kaya, E. M., Huang, N., & Elhilali, M. (2020). Pitch, Timbre and Intensity Interdependently Modulate Neural Responses to Salient Sounds. *Neuroscience*, *440*, 1–14. doi: 10.1016/j.neuroscience.2020.05.018

Kayser, H., Ewert, S. D., Anemüller, J., Rohdenburg, T., Hohmann, V., & Kollmeier, B. (2009). Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses. *Eurasip Journal on Advances in Signal Processing*, *2009*. doi: 10.1155/2009/298605

Keller, S., Tschan, F., Semmer, N. K., Holzer, E., Candinas, D., Brink, M., & Beldi, G. (2018). Noise in the Operating Room Distracts Members of the Surgical Team. An Observational Study. *World Journal of Surgery*, *42*(12), 3880–3887. doi: 10.1007/s00268-018-4730-7

Kern, M., Buia, A., Tonus, C., Weigel, T. F., Dittmar, R., Hanisch, E., & Zapf, D. (2019). Psychological stressors, resources and well-being of surgeons in Germany: A cross-sectional study. *Chirurg*, *90*(7), 576–584. doi: 10.1007/s00104-018-0780-5

Khalighinejad, B., Herrero, J. L., Mehta, A. D., & Mesgarani, N. (2019). Adaptation of the human auditory cortex to changing background noise. *Nature Communications*, *10*(1), 2509. doi: 10.1038/s41467-019-10611-4

Kidd, G., Jr., Mason, C. R., Best, V., & Marrone, N. (2010). Stimulus factors influencing spatial release from speech-on-speech masking. *The Journal of the Acoustical Society of America*, *128*(4), 1965–1978. doi: 10.1121/1.3478781

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception*, *36*(14), 1–16.

Klug, M., & Gramann, K. (2021). Identifying key factors for improving ICA-based decomposition of EEG data in mobile and stationary experiments. *European Journal of Neuroscience*, *54*(12), 8406–8420. doi: 10.1111/ejn.14992

Klug, M., Jeung, S., Wunderlich, A., Gehrke, L., Protzak, J., Djebbara, Z., ... Gramann, K. (2022). *The BeMoBIL Pipeline for automated analyses of multimodal mobile brain and body imaging data.* bioRxiv. doi: 10.1101/2022.09.29.510051

Kong, Y.-Y., Mullangi, A., & Ding, N. (2014). Differential modulation of auditory responses to attended and unattended speech in different listening conditions. *Hearing Research*, *316*, 73–81. doi: 10.1016/j.heares.2014.07.009

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. *Journal of Memory and Language*, *96*, 122–137. doi: 10.1016/j.jml.2017.05.005

Korte, S., Jaeger, M., Rosenkranz, M., & Bleichner, M. G. (2024). *From Beeps to Streets: Unveiling Sensory Input and Relevance Across Auditory Contexts.* bioRxiv. doi: 10.1101/2024.10.17.618668

Kothe, C., Shirazi, S. Y., Stenner, T., Medine, D., Boulay, C., Grivich, M. I., ... Makeig, S. (2024). *The Lab Streaming Layer for Synchronized Multimodal Recording* [Preprint]. bioRxiv. doi: 10.1101/2024.02.13.580071

Kounidas, G., Kastora, S., & Maini, S. K. (2022). Benefits of music on theatre staff in a paediatric theatre setting: A staff survey. *Irish Journal of Medical Science (1971 -)*. doi: 10.1007/s11845-021-02909-4

Kurmann, A., Peter, M., Tschan, F., Mühlemann, K., Candinas, D., & Beldi, G. (2011). Adverse effect of noise in the operating theatre on surgical-site infection. *British Journal of Surgery*, *98*(7), 1021–1025. doi: 10.1002/bjs.7496

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences*, *9*(2), 75–82. doi: 10.1016/j.tics.2004.12.004

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics*, *56*(2), 183–197. doi: 10.3758/BF03213897

Lebares, C. C., Guvva, E. V., Ascher, N. L., O'Sullivan, P. S., Harris, H. W., & Epel, E. S. (2018). Burnout and Stress Among US Surgery Residents: Psychological Distress and Resilience. *Journal of the American College of Surgeons*, *226*(1), 80–90. doi: 10.1016/j.jamcollsurg.2017

.10.010
Leitsmann, C., Uhlig, A., Popeneciu, I. V., Boos, M., Ahyai, S. A., Schmid, M., ... Friedrich, M. (2021). The Silent Operation Theatre Optimisation System (SOTOS©) to reduce noise pollution during da Vinci robot-assisted laparoscopic radical prostatectomy. *Journal of Robotic Surgery*, *15*(4), 519–527. doi: 10.1007/s11701-020-01135-x
Lijffijt, M., Lane, S. D., Meier, S. L., Boutros, N. N., Burroughs, S., Steinberg, J. L., ... Swann, A. C. (2009). P50, N100, and P200 sensory gating: Relationships with behavioral inhibition, attention, and working memory. *Psychophysiology*, *46*(5), 1059–1068. doi: 10.1111/j.1469-8986.2009.00845.x
Lim, C., Barragan, J. A., Farrow, J. M., Wachs, J. P., Sundaram, C. P., & Yu, D. (2023). Physiological Metrics of Surgical Difficulty and Multi-Task Requirement during Robotic Surgery Skills. *Sensors*, *23*(9), 4354. doi: 10.3390/s23094354
Little, J. S., Martin, F. H., & Thomson, R. H. S. (2010). Speech versus non-speech as irrelevant sound: Controlling acoustic variation. *Biological Psychology*, *85*(1), 62–70. doi: 10.1016/j.biopsycho.2010.05.004
Lopus, D. (2012). Sounds or Silence? Auditory Stimulus Effects on Fine Motor Skills. In *40th Annual Western Pennsylvania Undergraduate Psychology Conference* (p. 5).
Lorah, J. (2018). Effect size measures for multilevel models: Definition, interpretation, and TIMSS example. *Large-scale Assessments in Education*, *6*(1), 8. doi: 10.1186/s40536-018-0061-2
Lorenzi, C., Apoux, F., Grinfeder, E., Krause, B., Miller-Viacava, N., & Sueur, J. (2023). Human Auditory Ecology: Extending Hearing Research to the Perception of Natural Soundscapes by Humans in Rapidly Changing Environments. *Trends in Hearing*, *27*, 23312165231212032. doi: 10.1177/23312165231212032
Luck, S. (2014). *An introduction to the event-related potential technique*. MIT Press, Cambridge, Massachusetts.
Mahajan, Y., Kim, J., & Davis, C. (2020). Does working memory protect against auditory distraction in older adults? *BMC Geriatrics*, *20*(1), 515. doi: 10.1186/s12877-020-01909-w
Maier-Hein, L., Eisenmann, M., Sarikaya, D., März, K., Collins, T., Malpani, A., ... Speidel, S. (2022). Surgical data science – from concepts toward clinical translation. *Medical Image Analysis*, *76*, 102306. doi: 10.1016/j.media.2021.102306
Maimon, N. B., Bez, M., Drobot, D., Molcho, L., Intrator, N., Kakiashvilli, E., & Bickel, A. (2022). Continuous Monitoring of Mental Load During Virtual Simulator Training for Laparoscopic Surgery Reflects Laparoscopic Dexterity: A Comparative Study Using a Novel Wireless Device. *Frontiers in Neuroscience*, *15*.
Major, S., Carpenter, K., Beyer, L., Kwak, H., Dawson, G., & Murias, M. (2020). The Influence of Background Auditory Noise on P50 and N100 Suppression Elicited by the Paired-Click Paradigm. *Journal of Psychophysiology*, *34*(3), 171–178. doi: 10.1027/0269-8803/a000245
Marois, A., & Vachon, F. (2024). Psychophysiological Markers of Auditory Distraction: A Scoping Review. *Auditory Perception & Cognition*, *7*(1), 9–49. doi: 10.1080/25742442.2023.2274270
Marrelli, M., Gentile, S., Palmieri, F., Paduano, F., & Tatullo, M. (2014). Correlation between Surgeon's Experience, Surgery Complexity and the Alteration of Stress Related Physiological Parameters. *PLOS ONE*, *9*(11), e112444. doi: 10.1371/journal.pone.0112444
Marsh, J. E., Bell, R., Röer, J. P., & Hodgetts, H. M. (2024). Emerging perspectives on distraction and task interruptions: Metacognition, cognitive control and facilitation - part I. *Journal of Cognitive Psychology*, *36*(1), 1–7. doi: 10.1080/20445911.2024.2314974
Matusz, P. J., Dikker, S., Huth, A. G., & Perrodin, C. (2019). Are We Ready for Real-world Neuroscience? *Journal of Cognitive Neuroscience*, *31*(3), 327–338. doi: 10.1162/jocn_e_01276
May, P. J. C., & Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology*, *47*(1), 66–122. doi: 10.1111/j.1469-8986.2009.00856.x
McKendricka, R. D., & Cherry, E. (2018). A deeper look at the NASA TLX and where it falls

short. *Proceedings of the Human Factors and Ergonomics Society*, *1*, 44–48. doi: 10.1177/1541931218621010

Meiser, A., & Bleichner, M. G. (2022). Ear-EEG compares well to cap-EEG in recording auditory ERPs: A quantification of signal loss. *Journal of Neural Engineering*, *19*(2), 026042. doi: 10.1088/1741-2552/ac5fcb

Meiser, A., Tadel, F., Debener, S., & Bleichner, M. G. (2020). The Sensitivity of Ear-EEG: Evaluating the Source-Sensor Relationship Using Forward Modeling. *Brain Topography*, *33*(6), 665–676. doi: 10.1007/s10548-020-00793-2

Mentis, H. M., Chellali, A., Manser, K., Cao, C. G. L., & Schwaitzberg, S. D. (2016). A systematic review of the effect of distraction on surgeon performance: Directions for operating room policy and surgical training. *Surgical Endoscopy*, *30*(5), 1713–1724. doi: 10.1007/s00464-015-4443-z

Mesaros, A., Heittola, T., Eronen, A., & Virtanen, T. (2010). Acoustic event detection in real life recordings. In *2010 18th European Signal Processing Conference* (pp. 1267–1271).

Mesik, J., & Wojtczak, M. (2023). The effects of data quantity on performance of temporal response function analyses of natural speech processing. *Frontiers in Neuroscience*, *16*. doi: 10.3389/fnins.2022.963629

Miller, M. W., Rietschel, J. C., McDonald, C. G., & Hatfield, B. D. (2011). A novel approach to the physiological measurement of mental workload. *International Journal of Psychophysiology*, *80*(1), 75–78. doi: 10.1016/j.ijpsycho.2011.02.003

Mirkovic, B., Bleichner, M. G., De Vos, M., & Debener, S. (2016). Target speaker detection with concealed EEG around the ear. *Frontiers in Neuroscience*, *10*(JUL), 1–11. doi: 10.3389/fnins.2016.00349

Mirkovic, B., Debener, S., Jaeger, M., & De Vos, M. (2015). Decoding the attended speech stream with multi-channel EEG: Implications for online, daily-life applications. *Journal of Neural Engineering*, *12*(4), 46007. doi: 10.1088/1741-2560/12/4/046007

Mirkovic, B., Debener, S., Schmidt, J., Jaeger, M., & Neher, T. (2019). Effects of directional sound processing and listener's motivation on EEG responses to continuous noisy speech: Do normal-hearing and aided hearing-impaired listeners differ? *Hearing Research*, *377*, 260–270. doi: 10.1016/j.heares.2019.04.005

Modi, H. N., Singh, H., Darzi, A., & Leff, D. R. (2020). Multitasking and Time Pressure in the Operating Room: Impact on Surgeons' Brain Function. *Annals of Surgery*, *272*(4), 648–657. doi: 10.1097/SLA.0000000000004208

Molloy, K., Lavie, N., & Chait, M. (2019). Auditory Figure-Ground Segregation Is Impaired by High Visual Load. *Journal of Neuroscience*, *39*(9), 1699–1708. doi: 10.1523/JNEUROSCI.2518-18.2018

Moorthy, K., Munz, Y., Dosis, A., Bann, S., & Darzi, A. (2003). The effect of stress-inducing conditions on the performance of a laparoscopic task. *Surgical Endoscopy and Other Interventional Techniques*, *17*(9), 1481–1484. doi: 10.1007/s00464-002-9224-9

Moorthy, K., Munz, Y., Sarker, S. K., & Darzi, A. (2003). Objective assessment of technical skills in surgery. *British Medical Journal*, *327*(7422), 1032–1037. doi: 10.1136/bmj.327.7422.1032

Moorthy, K., Munz, Y., Undre, S., & Darzi, A. (2004). Objective evaluation of the effect of noise on the performance of a complex laparoscopic task. *Surgery*, *136*(1), 25–30. doi: 10.1016/j.surg.2003.12.011

Moray, N. (1959). Attention in Dichotic Listening: Affective Cues and the Influence of Instructions. *Quarterly Journal of Experimental Psychology*, *11*(1), 56–60. doi: 10.1080/17470215908416289

Mullen, T. R., Kothe, C. A., Chi, M., Ojeda, A., Kerth, T., Makeig, S., … Cauwenberghs, G. (2015). Real-time Neuroimaging and Cognitive Monitoring Using Wearable Dry EEG. *Physiology & behavior*, *176*(1), 100–106. doi: 10.1109/TBME.2015.2481482.

Murphy, G., Groeger, J. A., & Greene, C. M. (2016). Twenty years of load theory—Where are we now, and where should we go next? *Psychonomic Bulletin & Review*, *23*(5), 1316–1340. doi:

10.3758/s13423-015-0982-5
Murray, M. M., Brunet, D., & Michel, C. M. (2008). Topographic ERP Analyses: A Step-by-Step Tutorial Review. *Brain Topography*, *20*(4), 249–264. doi: 10.1007/s10548-008-0054-5
Näätänen, R. (1982). Processing Negativity: An Evoked-Potential Reflection of Selective Attention. *Psychological Bulletin*, *92*(3), 605–640. doi: 10.1037/0033-2909.92.3.605
Näätänen, R., & Picton, T. (1987). The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure. *Psychophysiology*, *24*(4), 375–425. doi: 10.1111/j.1469-8986.1987.tb00311.x
Oiamo, T. H., Luginaah, I. N., & Baxter, J. (2015, December). Cumulative effects of noise and odour annoyances on environmental and health related quality of life. *Social Science & Medicine*, *146*, 191–203. doi: 10.1016/j.socscimed.2015.10.043
O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., ... Lalor, E. C. (2015). Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cerebral Cortex*, *25*(7), 1697–1706. doi: 10.1093/cercor/bht355
Padmakumar, A. D., Cohen, O., Churton, A., Groves, J. B., Mitchell, D. A., & Brennan, P. A. (2017). Effect of noise on tasks in operating theatres: A survey of the perceptions of healthcare staff. *British Journal of Oral and Maxillofacial Surgery*, *55*(2), 164–167. doi: 10.1016/j.bjoms.2016.10.011
Parasuraman, R. (2003). Neuroergonomics: Research and practice. *Theoretical Issues in Ergonomics Science*, *4*(1-2), 5–20. doi: 10.1080/14639220210199753
Patel, S. H., & Azzam, P. N. (2005). Characterization of N200 and P300: Selected Studies of the Event-Related Potential. *International Journal of Medical Sciences*, *2*(4), 147–154.
Persoon, M. C., Broos, H. J. H. P., Witjes, J. A., Hendrikx, Ad. J. M., & Scherpbier, A. J. J. M. (2011). The effect of distractions in the operating room during endourological procedures. *Surgical Endoscopy*, *25*(2), 437–443. doi: 10.1007/s00464-010-1186-8
Peryer, G., Noyes, J., Pleydell-Pearce, K., & Lieven, N. (2005). Auditory Alert Characteristics: A Survey of Pilot Views. *The International Journal of Aviation Psychology*, *15*(3), 233–250. doi: 10.1207/s15327108ijap1503_2
Petersen, E. B., Wöstmann, M., Obleser, J., & Lunner, T. (2017). Neural tracking of attended versus ignored speech is differentially affected by hearing loss. *Journal of Neurophysiology*, *117*(1), 18–27. doi: 10.1152/jn.00527.2016
Picton, T. W., Alain, C., Otten, L., Ritter, W., & Achim, A. (2000). Mismatch negativity: Different water in the same river. *Audiology & Neuro-Otology*, *5*(3-4), 111–139.
Picton, T. W., & Hillyard, S. A. (1974). Human auditory evoked potentials. II: Effects of attention. *Electroencephalography and Clinical Neurophysiology*, *36*, 191–200. doi: 10.1016/0013-4694(74)90156-4
Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, *198*, 181–197. doi: 10.1016/j.neuroimage.2019.05.026
Plapous, C., Marro, C., & Scalart, P. (2006). Improved Signal-to-Noise Ratio Estimation for Speech Enhancement. *IEEE Transactions on Audio, Speech, and Language Processing*, *14*(6), 2098–2108. doi: 10.1109/TASL.2006.872621
Pleban, D., Radosz, J., Kryst, L., & Surgiewicz, J. (2021). Assessment of working conditions in medical facilities due to noise. *International Journal of Occupational Safety and Ergonomics*, *27*(4), 1199–1206. doi: 10.1080/10803548.2021.1987692
Pluyter, J. R., Buzink, S. N., Rutkowski, A.-F., & Jakimowicz, J. J. (2010). Do absorption and realistic distraction influence performance of component task surgical procedure? *Surgical Endoscopy*, *24*(4), 902–907. doi: 10.1007/s00464-009-0689-7
Pohlhausen, J., Nespoli, F., & Bitzer, J. (2024). Enhancing Speech Privacy with LPC Modifications. In *Proc. SPSC 2024* (pp. 80–85). doi: 10.21437/SPSC.2024-14
Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. doi: 10.1016/j.clinph.2007.04.019

Poolton, J. M., Zhu, F. F., Malhotra, N., Leung, G. K. K., Fan, J. K. M., & Masters, R. S. W. (2016). Multitask training promotes automaticity of a fundamental laparoscopic skill without compromising the rate of skill learning. *Surgical Endoscopy*, *30*(9), 4011–4018. doi: 10.1007/s00464-015-4713-9

Price, C., Thierry, G., & Griffiths, T. (2005). Speech-specific auditory processing: Where is it? *Trends in Cognitive Sciences*, *9*(6), 271–276. doi: 10.1016/j.tics.2005.03.009

Protzak, J., & Gramann, K. (2018). Investigating established EEG parameter during real-world driving. *Frontiers in Psychology*, *9*(NOV), 1–11. doi: 10.3389/fpsyg.2018.02289

Pugh, C. M., Ghazi, A., Stefanidis, D., Schwaitzberg, S. D., Martino, M. A., & Levy, J. S. (2020). How Wearable Technology Can Facilitate AI Analysis of Surgical Videos. *Annals of Surgery Open*, *1*(2), e011. doi: 10.1097/AS9.0000000000000011

Regenbogen, C., Vos, M. D., Debener, S., Turetsky, B. I., Mößnang, C., Finkelmeyer, A., … Kellermann, T. (2012). Auditory Processing under Cross-Modal Visual Load Investigated with Simultaneous EEG-fMRI. *PLOS ONE*, *7*(12), e52267. doi: 10.1371/journal.pone.0052267

Rennies, J., Ohlenbusch, M., Volgenandt, A., Spitz, T., Baumgartner, H., Rollwage, C., … Weber, V. (2023). Analyse und algorithmische Optimierung von Geräuschkulissen und Sprachkommunikation im OP- Saal. In *Fortschritte der akustik – daga 2023* (pp. 646–649).

Rieger, A., Stoll, R., Kreuzfeld, S., Behrens, K., & Weippert, M. (2014). Heart rate and heart rate variability as indirect markers of surgeons' intraoperative stress. *International Archives of Occupational and Environmental Health*, *87*(2), 165–174. doi: 10.1007/s00420-013-0847-z

Rosburg, T., Trautner, P., Elger, C. E., & Kurthen, M. (2009). Attention effects on sensory gating — Intracranial and scalp recordings. *NeuroImage*, *48*(3), 554–563. doi: 10.1016/j.neuroimage.2009.06.063

Rosenkranz, M., & Bleichner, M. (2022, October). *Data and experiment from: Investigating the attentional focus to workplace-related soundscapes in a complex audio-visual-motor task using EEG.* Zenodo. Retrieved from https://doi.org/10.5281/zenodo.7147701 doi: 10.5281/zenodo.7147701

Rosenkranz, M., Cetin, T., Uslar, V. N., & Bleichner, M. G. (2023). Investigating the attentional focus to workplace-related soundscapes in a complex audio-visual-motor task using EEG. *Frontiers in Neuroergonomics*, *3*.

Rosenkranz, M., Haupt, T., Jaeger, M., Uslar, V. N., & Bleichner, M. G. (2024). Using mobile EEG to study auditory work strain during simulated surgical procedures. *Scientific Reports*, *14*(1), 24026. doi: 10.1038/s41598-024-74946-9

Ross, B., & Tremblay, K. (2009). Stimulus experience modifies auditory neuromagnetic responses in young and older listeners. *Hearing research*, *248*(1-2), 48–59. doi: 10.1016/j.heares.2008.11.012

Roye, A., Jacobsen, T., & Schröger, E. (2013). Discrimination of personally significant from nonsignificant sounds: A training study. *Cognitive, Affective, & Behavioral Neuroscience*, *13*(4), 930–943. doi: 10.3758/s13415-013-0173-7

SanMiguel, I., Corral, M.-J., & Escera, C. (2008). When Loading Working Memory Reduces Distraction: Behavioral and Electrophysiological Evidence from an Auditory-Visual Distraction Paradigm. *Journal of Cognitive Neuroscience*, *20*(7), 1131–1145. doi: 10.1162/jocn.2008.20078

Scalart, P. (2023). *Wiener filter for noise reduction and speech enhancement.* https://www.mathworks.com/matlabcentral/fileexchange/24462-wiener-filter-for-noise-reduction-and-speech-enhancement. ([Online; accessed April, 2023])

Scanlon, J. E., Townsend, K. A., Cormier, D. L., Kuziek, J. W., & Mathewson, K. E. (2019). Taking off the training wheels: Measuring auditory P3 during outdoor cycling using an active wet EEG system. *Brain Research*, *1716*, 50–61. doi: 10.1016/j.brainres.2017.12.010

Scerbo, M. W., Britt, R. C., & Stefanidis, D. (2017). Differences in mental workload between traditional and single-incision laparoscopic procedures measured with a secondary task. *The American Journal of Surgery*, *213*(2), 244–248. doi: 10.1016/j.amjsurg.2016.09.056

Schlickum, M., Hedman, L., Enochsson, L., Henningsohn, L., Kjellin, A., & Felländer-Tsai, L. (2011). Surgical Simulation Tasks Challenge Visual Working Memory and Visual-Spatial Ability Differently. *World Journal of Surgery*, *35*(4), 710–715. doi: 10.1007/s00268-011-0981-2

Schlossmacher, I., Dellert, T., Bruchmann, M., & Straube, T. (2021). Dissociating neural correlates of consciousness and task relevance during auditory processing. *NeuroImage*, *228*, 117712. doi: 10.1016/j.neuroimage.2020.117712

Schröger, E. (1997). On the detection of auditory deviations: A pre-attentive activation model. *Psychophysiology*, *34*(3), 245–257. doi: 10.1111/j.1469-8986.1997.tb02395.x

Schutte, M., Marks, A., Wenning, E., & Griefahn, B. (2007). The development of the noise sensitivity questionnaire Development of the NoiSeQ. *Noise and Health*, *9*(34), 15–24. doi: 10.4103/1463-1741.34700

Sevdalis, N., Healey, A. N., & Vincent, C. A. (2007). Distracting communications in the operating theatre. *Journal of Evaluation in Clinical Practice*, *13*(3), 390–394. doi: 10.1111/j.1365-2753.2006.00712.x

Sevdalis, N., Undre, S., McDermott, J., Giddie, J., Diner, L., & Smith, G. (2014). Impact of Intraoperative Distractions on Patient Safety: A Prospective Descriptive Study Using Validated Instruments. *World Journal of Surgery*, *38*(4), 751–758. doi: 10.1007/s00268-013-2315-z

Shafiei, S. B., Jing, Z., Attwood, K., Iqbal, U., Arman, S., Hussein, A. A., … Guru, K. (2021). Association between Functional Brain Network Metrics and Surgeon Performance and Distraction in the Operating Room. *Brain Sciences*, *11*(4), 468. doi: 10.3390/brainsci11040468

Shanafelt, T. D., Hasan, O., Dyrbye, L. N., Sinsky, C., Satele, D., Sloan, J., & West, C. P. (2015). Changes in Burnout and Satisfaction With Work-Life Balance in Physicians and the General US Working Population Between 2011 and 2014. *Mayo Clinic Proceedings*, *90*(12), 1600–1613. doi: 10.1016/j.mayocp.2015.08.023

Shapiro, R. A., & Berland, T. (1972). Noise in the Operating Room. *New England Journal of Medicine*, *287*(24), 1236–1238. doi: 10.1056/NEJM197212142872407

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, *12*(5), 182–186. doi: 10.1016/j.tics.2008.02.003

Shinn-Cunningham, B. G., Mehraei, G., Bressler, S., & Masud, S. (2013). Influences of perceptual continuity on everyday listening. *Proceedings of Meetings on Acoustics*, *19*(1), 010026. doi: 10.1121/1.4799260

Siu, K.-C., Suh, I. H., Mukherjee, M., Oleynikov, D., & Stergiou, N. (2010). The impact of environmental noise on robot-assisted laparoscopic surgical performance. *Surgery*, *147*(1), 107–113. doi: 10.1016/j.surg.2009.08.010

Sörqvist, P. (2015). On interpretation and task selection: The sub-component hypothesis of cognitive noise effects. *Frontiers in Psychology*, *5*. doi: 10.3389/fpsyg.2014.01598

Sörqvist, P., Dahlström, Ö., Karlsson, T., & Rönnberg, J. (2016). Concentration: The Neural Underpinnings of How Cognitive Load Shields Against Distraction. *Frontiers in Human Neuroscience*, *10*.

Sörqvist, P., Stenfelt, S., & Rönnberg, J. (2012). Working Memory Capacity and Visual–Verbal Cognitive Load Modulate Auditory–Sensory Gating in the Brainstem: Toward a Unified View of Attention. *Journal of Cognitive Neuroscience*, *24*(11), 2147–2154. doi: 10.1162/jocn_a_00275

Straetmans, L., Adiloglu, K., & Debener, S. (2024). Neural speech tracking and auditory attention decoding in everyday life. *Frontiers in Human Neuroscience*, *18*. doi: 10.3389/fnhum.2024.1483024

Straetmans, L., Holtze, B., Debener, S., Jaeger, M., & Mirkovic, B. (2021). Neural tracking to go: Auditory attention decoding and saliency detection with mobile EEG. *Journal of Neural Engineering*, *18*(6), 066054. doi: 10.1088/1741-2552/ac42b5

Suárez, J. X., Gramann, K., Ochoa, J. F., Toro, J. P., Mejía, A. M., & Hernández, A. M. (2022). Changes in brain activity of trainees during laparoscopic surgical virtual training assessed with electroencephalography. *Brain Research*, *1783*, 147836. doi: 10.1016/j.brainres.2022

.147836
Suh, I. H., Chien, J.-H., Mukherjee, M., Park, S.-H., Oleynikov, D., & Siu, K.-C. (2010). The negative effect of distraction on performance of robot-assisted surgical skills in medical students and residents. *The International Journal of Medical Robotics and Computer Assisted Surgery*, *6*(4), 377–381. doi: 10.1002/rcs.338
Suh, I. H., LaGrange, C. A., Oleynikov, D., & Siu, K.-C. (2016). Evaluating Robotic Surgical Skills Performance Under Distractive Environment Using Objective and Subjective Measures. *Surgical Innovation*, *23*(1), 78–89. doi: 10.1177/1553350615596637
Szafranski, C., Kahol, K., Ghaemmaghami, V., Smith, M., & Ferrara, J. J. (2009). Distractions and surgical proficiency: An educational perspective. *The American Journal of Surgery*, *198*(6), 804–810. doi: 10.1016/j.amjsurg.2009.04.027
Szalma, J. L., & Hancock, P. A. (2011). Noise effects on human performance: A meta-analytic synthesis. *Psychological Bulletin*, *137*(4), 682–707. doi: 10.1037/a0023987
Thomaschewski, M., Heldmann, M., Uter, J. C., Varbelow, D., Münte, T. F., & Keck, T. (2021). Changes in attentional resources during the acquisition of laparoscopic surgical skills. *BJS Open*, *5*(2), zraa012. doi: 10.1093/bjsopen/zraa012
Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological Review*, *76*(3), 282–299. doi: 10.1037/h0027242
Tsiou, C., Efthymiatos, G., & Katostaras, T. (2008). Noise in the operating rooms of Greek hospitals. *The Journal of the Acoustical Society of America*, *123*(2), 757–765. doi: 10.1121/1.2821972
Ueda, K., Nakajima, Y., Kattner, F., & Ellermeier, W. (2019). Irrelevant speech effects with locally time-reversed speech: Native vs non-native languagea). *The Journal of the Acoustical Society of America*, *145*(6), 3686–3694. doi: 10.1121/1.5112774
Ullsperger, P., Erdmann, U., Freude, G., & Dehoff, W. (2006). When sound and picture do not fit: Mismatch negativity and sensory interaction. *International Journal of Psychophysiology*, *59*(1), 3–7. doi: 10.1016/j.ijpsycho.2005.06.007
van Harten, A., Gooszen, H. G., Koksma, J. J., Niessen, T. J., & Abma, T. A. (2021). An observational study of distractions in the operating theatre. *Anaesthesia*, *76*(3), 346–356. doi: 10.1111/anae.15217
Vanthornhout, J., Decruy, L., & Francart, T. (2019). Effect of Task and Attention on Neural Tracking of Speech. *Frontiers in Neuroscience*, *13*(September), 1–11. doi: 10.3389/fnins.2019.00977
Verkuilen, J., & Smithson, M. (2012). Mixed and Mixture Regression Models for Continuous Bounded Responses Using the Beta Distribution. *Journal of Educational and Behavioral Statistics*, *37*(1), 82–113. doi: 10.3102/1076998610396895
Vigliocco, G., Convertino, L., De Felice, S., Gregorians, L., Kewenig, V., Mueller, M. A. E., ... Spiers, H. J. (2024). Ecological brain: Reframing the study of human behaviour and cognition. *Royal Society Open Science*, *11*(11), 240762. doi: 10.1098/rsos.240762
Vogel, M., Braungardt, T., Meyer, W., & Schneider, W. (2012). The effects of shift work on physical and mental health. *Journal of Neural Transmission*, *119*(10), 1121–1132. doi: 10.1007/s00702-012-0800-4
Volpert-Esmond, H. I., Merkle, E. C., Levsen, M. P., Ito, T. A., & Bartholow, B. D. (2018). Using trial-level data and multilevel modeling to investigate within-task change in event-related potentials. *Psychophysiology*, *55*(5), e13044. doi: 10.1111/psyp.13044
Volpert-Esmond, H. I., Page-Gould, E., & Bartholow, B. D. (2021). Using multilevel models for the analysis of event-related potentials. *International Journal of Psychophysiology*, *162*, 145–156. doi: 10.1016/j.ijpsycho.2021.02.006
Wascher, E., Heppner, H., & Hoffmann, S. (2014). Towards the measurement of event-related EEG activity in real-life working environments. *International Journal of Psychophysiology*, *91*(1), 3–9. doi: 10.1016/j.ijpsycho.2013.10.006
Wascher, E., Reiser, J., Rinkenauer, G., Larrá, M., Dreger, F. A., Schneider, D., ... Arnau, S. (2021). Neuroergonomics on the Go: An Evaluation of the Potential of Mobile EEG for Workplace Assessment and Design. *Human Factors*, *00*(0), 1–21. doi: 10.1177/00187208211007707

Way, T. J., Long, A., Weihing, J., Ritchie, R., Jones, R., Bush, M., & Shinn, J. B. (2013). Effect of noise on auditory processing in the operating room. *Journal of the American College of Surgeons*, *216*(5), 933–938. doi: 10.1016/j.jamcollsurg.2012.12.048

Weigl, M., Antoniadis, S., Chiapponi, C., Bruns, C., & Sevdalis, N. (2015). The impact of intra-operative interruptions on surgeons' perceived workload: An observational study in elective general and orthopedic surgery. *Surgical Endoscopy*, *29*(1), 145–153. doi: 10.1007/s00464-014-3668-6

Wheelock, A., Suliman, A., Wharton, R., Babu, E. D., Hull, L., Vincent, C., ... Arora, S. (2015). The Impact of Operating Room Distractions on Stress, Workload, and Teamwork. *Annals of Surgery*, *261*(6), 1079. doi: 10.1097/SLA.0000000000001051

Wickens, C. D. (2008). Multiple Resources and Mental Workload. *Human Factors*, *50*(3), 449–455. doi: 10.1518/001872008X288394

Widmer, L. W., Keller, S., Tschan, F., Semmer, N. K., Holzer, E., Candinas, D., & Beldi, G. (2018). More Than Talking About the Weekend: Content of Case-Irrelevant Communication Within the OR Team. *World Journal of Surgery*, *42*(7), 1. doi: 10.1007/s00268-017-4442-4

Wiedenmann, E., Daeglau, M., Otten, J., Grimm, G., Hohmann, V., & Debener, S. (2023). The Influence of Likeability Ratings of Audio-Visual Stimuli on Cortical Speech Tracking with Mobile EEG in Virtual Environments. In *DAGA 2023 Hamburg* (Vol. 49).

Wilson, M. R., Poolton, J. M., Malhotra, N., Ngo, K., Bright, E., & Masters, R. S. W. (2011). Development and Validation of a Surgical Workload Measure: The Surgery Task Load Index (SURG-TLX). *World Journal of Surgery*, *35*(9), 1961–1969. doi: 10.1007/s00268-011-1141-4

Wong, B. W. L., Huo, S., & Maurer, U. (2023). *Adaptation and Mismatch Negativity (MMN): Comparing First vs. Subsequent Repetitions in a Roving EEG Paradigm with Minimized Expectations* [Preprint]. Preprints. doi: 10.22541/au.168323405.54729335/v1

Xie, Z., Brodbeck, C., & Chandrasekaran, B. (2023). Cortical Tracking of Continuous Speech Under Bimodal Divided Attention. *Neurobiology of Language*, *4*(2), 318–343. doi: 10.1162/nol_a_00100

Yang, C., Heinze, J., Helmert, J., Weitz, J., Reissfelder, C., & Mees, S. T. (2017). Impaired laparoscopic performance of novice surgeons due to phone call distraction: A single-centre, prospective study. *Surgical Endoscopy*, *31*(12), 5312–5317. doi: 10.1007/s00464-017-5609-7

You, S., Xu, F., Wu, Y., Qin, S., Shu, B., Chen, Y., ... Duan, G. (2024). Effect of noise isolation using noise-cancelling headphones during laparoscopic surgery for postoperative pain reduction: A randomized clinical trial. *Journal of Clinical Anesthesia*, *92*, 111286. doi: 10.1016/j.jclinane.2023.111286

Zakeri, Z., Mansfield, N., Sunderland, C., & Omurtag, A. (2020). Physiological correlates of cognitive load in laparoscopic surgery. *Scientific Reports*, *10*(1), 12927. doi: 10.1038/s41598-020-69553-3

Zander, T. O., Shetty, K., Lorenz, R., Leff, D. R., Krol, L. R., Darzi, A. W., ... Yang, G.-Z. (2017). Automated Task Load Detection with Electroencephalography: Towards Passive Brain–Computer Interfacing in Robotic Surgery. *Journal of Medical Robotics Research*, *02*(01), 1750003. doi: 10.1142/s2424905x17500039

Zhang, Y., Hedo, R., Rivera, A., Rull, R., Richardson, S., & Tu, X. M. (2019). Post hoc power analysis: Is it an informative and meaningful analysis? *General Psychiatry*, *32*(4), e100069. doi: 10.1136/gpsych-2019-100069

Zuk, N. J., Murphy, J. W., Reilly, R. B., & Lalor, E. C. (2021). Envelope reconstruction of speech and music highlights stronger tracking of speech at low frequencies. *PLOS Computational Biology*, *17*(9), e1009358. doi: 10.1371/journal.pcbi.1009358

Declaration on the use of generative AI

In the preparation of this thesis, I have used the following generative AI based systems:

ChatGPT 4
DeepL Write

I declare that I:

have used the above-mentioned AI systems for spell-checking, grammar, and wording of this thesis,

have read the "Recommendations for the use of generative AI in doctoral studies, especially dissertations, at the University of Oldenburg" (12th November 2024),

have actively informed myself about the capabilities and limitations of the above-mentioned AI systems to the extent that I can use them responsibly,

have verified that the content generated by the above-mentioned AI systems and adopted by me is factually correct,

am aware that, as the author of this work, I am responsible for the information and statements made in it.

Oldenburg, February 19, 2025

Marc Rosenkranz

Declaration I

I hereby confirm that Marc Rosenkranz contributed to the aforementioned studies as stated below:

Study I: Rosenkranz M., Cetin T., Uslar V.N. and Bleichner M.G. (2023). Investigating the attentional focus to workplace-related soundscapes in a complex audio-visual-motor task using EEG. *Front. Neuroergon.*. 3:1062227. doi: 10.3389/fnrgo.2022.1062227

Author contributions: Author contributions: MR and MB conceptualized the experiment. MR performed the data acquisition, analyzed the data, and wrote the manuscript to which TC, VU, and MB contributed with critical revisions. All authors approved the final version and agreed to be accountable for this work.

Study II: Rosenkranz, M., Haupt, T., Jaeger, M., Uslar, V.N., & Bleichner, M. G. (2024). Using mobile EEG to study auditory work strain during simulated surgical procedures. *Sci Rep*, 3. https://doi.org/10.1038/s41598-024-74946-9

Author contributions: M.R. and M.B. conceptualized the experiment. M.R. performed the data acquisition and analyzed the data, which was assisted by T.H., M.J., and M.B.. M.R. wrote the manuscript to which T.H., M.J., V.U., and MB contributed with critical revisions. All authors reviewed the manuscript.

Study III: Rosenkranz, M., Uslar, V.N., Weye, D. & Bleichner, M. G. (2025, preprint). The effect of task demand on EEG responses to irrelevant sound and speech in simulated surgical environments. *bioRxiv*. https://doi.org/10.1101/2025.02.13.638036

Author contributions: MR and MGB conceptualized the experiment. MR performed the data acquisition and analyzed the data. MR wrote the manuscript to which MGB, VNU, and DW contributed with critical revisions. All authors reviewed the manuscript.

Oldenburg, February 19, 2025

Martin G. Bleichner

Declaration II

I have completed the work independently and used the indicated facilities.

This dissertation is my own work. All the sources of information have been acknowledged by means of references.

This dissertation has neither as a whole nor in part been published or submitted to assessment in a doctoral procedure at another university.

This is to confirm that I am aware of the guidelines of good scientific practice of the Carl von Ossietzky University of Oldenburg and that I observed them.

This is to confirm that I have not availed myself of any commercial placement or consulting services in connection with my promotion procedure.

Oldenburg, February 19, 2025

Marc Rosenkranz

This thesis was typeset with $\LaTeX\,2_\varepsilon$. It uses the *Clean Thesis* style developed by Ricardo Langner. The design of the *Clean Thesis* style is inspired by user guide documents from Apple Inc.

Download the *Clean Thesis* style at http://cleanthesis.der-ric.de/.