THESIS FOR THE DEGREE OF LICENTIATE OF ENGINEERING

Implications of Speech Exposure for Working Memory and Perceived Workload

ELIN HEDLUND

Department of Architecture and Civil Engineering
Division of Applied Acoustics
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden, 2026

Implications of Speech Exposure for Working Memory and Perceived Workload

ELIN HEDLUND

Technical report: 2026-10
Lic/Architecture and Civil Engineering/Chalmers University of Technology

Department of Architecture and Civil Engineering
Division of Applied Acoustics
Chalmers University of Technology
SE-412 96 Göteborg,
Sweden
Phone: +46(0)31 772 1000

Cover:
Time-frequency representation (spectrogram) of a speech signal, illustrating how its spectral content evolves over time.

Printed by Chalmers Digitaltryck,
Gothenburg, Sweden 2026.

Implications of Speech Exposure for Working Memory and Perceived Workload

ELIN HEDLUND

Department of Architecture and Civil Engineering
Division of Applied Acoustics
CHALMERS UNIVERSITY OF TECHNOLOGY

Abstract

Background speech is common in many modern work environments, particularly in open-plan offices, where it has been associated with increased disturbance and impaired cognitive performance. Although the effects of irrelevant speech on cognition have been demonstrated previously, the acoustic factors and cognitive processes contributing to these effects are still not fully understood.

This thesis investigates how background speech influences working memory performance and perceived workload under controlled acoustic conditions. A series of listening experiments was conducted in which participants performed cognitive tasks while being exposed to different speech conditions. The experiments examined the effects of varying reverberation time, semantic content, and sound level.

The results show that exposure to background speech negatively affects working memory performance and increases perceived workload compared to silence, and demonstrate that the disruptive effect of speech depends on characteristics of the signal itself. Higher speech levels were associated with stronger disruption, while speech-like temporal structure appeared to contribute more strongly to the effect than semantic content alone. Furthermore, perceived workload increased even when objective performance remained relatively stable, suggesting that maintaining performance under speech exposure may require additional mental effort.

The findings contribute to a broader understanding of how different characteristics of speech influence both cognitive performance and subjective experience. The work highlights the importance of considering both objective and subjective measures when evaluating the impact of speech exposure and suggest that reducing the prominence of background speech may be important for supporting cognitive functioning and limiting mental strain during sustained work tasks.

Keywords

speech exposure, working memory, perceived workload, listening experiment

List of Publications

Appended publications

This thesis is based on the following publications:

Paper I	**E. Hedlund**, L. Müller, W. Kropp, Influence of Speech Exposure on Working Memory: Exploring Reverberation Time in Open Plan Office Environments, *in Fortschritte der Akustik - DAGA 2024: 50 Jahrestagung für Akustik,* pp. 722-725, Hannover, Germany, 2024.
Paper II	**E. Hedlund**, L. Lindel, W. Kropp, Influence of Speech Characteristics on Working Memory and Perceived Workload: Findings from Experimental Listening Experiments, *Under review,* 2026.

Acknowledgments

Laura, thank you for your dedication and contribution to the progress of this work. Thank you to my supervisors, Wolfgang and Jens. Wolfgang, thank you for your guidance, support, and insightful feedback throughout this work. Jens, thank you for your help and encouragement during this licentiate.

Astrid, Christopher, Fati, Guillem, Hannes, Jannik, Jens, Krister, Leon, Monica, Patrik, and Tommi, thank you for your input and for the discussions we shared at Applied Acoustics.

The work presented in this thesis was performed at the Division of Applied Acoustics at Chalmers University of Technology within the project *Perception-based planning of the acoustic conditions in office environments using a virtual environment* [1]. The project was funded by Afa Försäkringar with project number 190278.

Friends and family, thank you for your patience, love, and encouragement.

Calle and Uno, I love you.

Contents

Part I

Summary

Chapter 1

Introduction

1.1 Speech Exposure in Modern Work Environments

Open-plan and activity-based office environments have become increasingly common in modern workplaces. In Sweden, more than 40% of office workers in 2024 were based in open-plan or activity-based offices [2]. This development has been driven by both technological and organizational changes. The widespread use of laptops and other mobile devices has reduced the need for fixed workstations, allowing employees to work more flexibly within shared environments. In addition, shifts in work practices, particularly following increased remote work, have further contributed to the implementation of flexible office layouts designed to accommodate varying levels of occupancy.

From an organizational perspective, open-plan offices are often motivated by considerations of efficiency and cost. These layouts allow for higher occupancy density and greater flexibility in space utilization compared to traditional

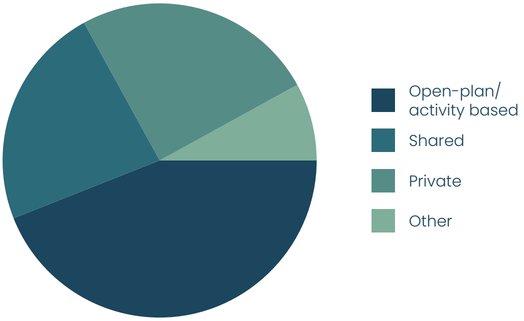

Figure 1.1: Distribution of office spaces types among employees working two or more days per week in Sweden in 2024 [2].

cellular offices. While such environments can facilitate communication and collaboration, they also introduce new challenges related to the acoustic environment.

A key concern in open-plan offices is the presence of background noise, where conversational speech has repeatedly been identified as one of the most disruptive components [3], [4], [5]. Although some occupants report positive aspects of open environments, a substantial body of research indicates that background speech can impair concentration, reduce task performance, and increase perceived disturbance and stress. These effects are particularly relevant for tasks that rely on sustained attention and working memory, where information must be temporarily stored and actively processed over short periods of time.

One explanation for this disruption is that speech, unlike many other environmental sounds, contains meaningful and dynamically changing information. Even when it is task-irrelevant, speech can involuntarily capture attention and interfere with cognitive processes involved in maintaining ordered representation in working memory. This type of interference has been extensively studied in the context of the irrelevant speech effect, where background speech disrupts performance in tasks requiring serial recall [6], [7]. Importantly, this suggests that the impact of speech is not solely dependent on its semantic content, but also on its temporal and spectral characteristics, which are likely to compete with internal cognitive processes.

The detrimental impact of speech is therefore not only related to its presence as a sound source, but also to its specific acoustic and cognitive properties [8], [9]. Speech signals are characterized by fluctuations in amplitude, frequency content, and rhythm, all of which can contribute to their ability to attract attention and interfere with task performance. However, it remains unclear which aspects of speech are most critical for this disruption. Factors such as speech intelligibility, sound level, and acoustic environment, including reverberation time, could all influence the degree to which speech affects cognitive performance, but their relative contributions are not fully understood.

This uncertainty also complicates the interpretation of acoustic descriptors commonly used in office acoustics. Measures intended to characterize speech transmission or acoustic quality often combine several underlying acoustic properties into single values. For example, the Speech Transmission Index (STI) reflects combined effects of factors including speech level, reverberation, and background noise on speech transmission. While such descriptors are useful for characterizing acoustic environments, it can be difficult to determine which specific properties of the sound environment are most strongly related to cognitive disruption.

In practice, acoustic interventions in office environments often focus on modifying the physical environment, for example, through the use of sound-absorbing materials. Although such measures can alter the acoustic properties of a space, their effectiveness in reducing cognitive disruption is still uncertain. This raises the question whether the primary driver of distraction lies in the acoustic environment itself or in specific characteristics of the speech signals.

1.2 Problem Formulation

Speech exposure has been shown to impair working memory and increase perceived workload, effects that are highly relevant in real work environments where background speech is common.

While these effects are well established, the mechanisms underlying the disruption are inherently multifaceted. Both the acoustic environment, such as reverberation time, and properties of the speech signal itself, including intelligibility, sound level, and temporal structure, can influence how speech is perceived and processed during cognitive tasks.

These factors are closely interrelated. For example, changes in sound level can alter the perceptual prominence of speech, while reverberation can affect both its intelligibility and temporal structure. As a result, the observed cognitive effects of speech exposure may reflect a combination of interacting influences rather than a single dominant factor. This makes it challenging to attribute changes in performance or perceived workload to a specific aspect of an acoustic conditions.

To better understand these relationships, it is necessary to examine the effects of individual factors under controlled conditions. By systematically varying characteristics of the speech signal and the acoustic environment, it becomes possible to disentangle their contributions and assess how they influence cognitive performance and perceived workload.

Against this background, the present work investigates how specific acoustic factors of speech exposure affect working memory and perceived workload.

1.3 Aim

This thesis aims to investigate how specific acoustic characteristics of speech exposure influence working memory performance and perceived workload. In particular, the work focuses on isolating individual factors related to both the speech signal and the acoustic environment. The investigated factors are

REVERBERATION TIME	Examining how variations in room acoustics, represented by different reverberation times, influence cognitive performance and perceived workload during speech exposure.
SEMANTIC CONTENT	Assessing the effect of linguistic information by comparing intelligible speech with speech signals lacking semantic content while preserving speech-like temporal and spectral characteristics.
SOUND LEVEL	Investigating how variations in sound level, and thereby perceptual prominence, influence the disruptive effect of speech on working memory and perceived workload.

Together, these factors represent complementary aspects of speech exposure, spanning both environmental and signal-related characteristics. Examining them under controlled conditions provides a basis for relating observed cognitive effects to specific acoustic properties, rather than to speech exposure as a general phenomenon.

1.4 Scope and Delimitations

This thesis is based on controlled listening experiments in which participants are exposed to speech under systematically varied acoustic conditions. The work focuses on isolating individual variables related to both the speech signal and the acoustic environment in order to examine their effects on cognitive performance and perceived workload.

The experimental approach is deliberately constrained to allow controlled manipulation of specific factors. As such, the study does not aim to replicate the full complexity of real-world environments, but rather to investigate selected aspects of speech exposure under well-defined conditions. This approach enables a systematic examination of the relationship between acoustic factors and cognitive responses, while maintaining a controlled and reproducible experimental framework.

Cognitive performance is assessed using the Operation Span Task (OSPAN) [10], which measures working memory capacity in tasks requiring simultaneous processing and storage. Perceived workload is evaluated using the NASA Task Load Index (NASA-TLX) [11], providing a subjective assessment of task demands under different acoustic conditions.

1.5 Structure of the Thesis

Following the current chapter, Part I is structured as follows.

CHAPTER 2 provides an overview of the theoretical and empirical background relevant to this thesis. Key concepts are introduced, related to speech exposure, working memory, and perceived workload, and previous research on how irrelevant speech and its acoustic characteristics influence cognitive performance and listening effort is reviewed.

CHAPTER 3 presents the acoustic and experimental framework underlying the work. It outlines how the acoustic environments were modeled and simulated, how the speech stimuli were constructed and manipulated, and how the listening experiments were designed, including the cognitive and subjective measures used to assess performance and workload.

CHAPTER 4 includes summaries of the appended papers, highlighting their individual problems, contributions and methodologies.

CHAPTER 5 discusses the findings of the included papers in relation to previous research and theoretical perspectives, and presents the overall conclusions that can be drawn from the thesis.

CHAPTER 6 outlines possible directions for future research, including inter-individual variability in susceptibility to speech distraction and questions related to ecological validity.

Part II contains the two appended papers.

Chapter 2

Auditory Processing and Cognitive Performance

When examining how individuals are affected by speech exposure in terms of cognitive processing and perceived workload, it is necessary to consider both auditory processing and the cognitive mechanisms involved in interpreting such input. Research in these areas has addressed a range of perspectives, including models of working memory, mechanisms of speech perception and the interaction between auditory input and task performance.

This chapter provides a conceptual background based on previous research relevant to this thesis. It introduces key concepts related to working memory, speech exposure, and cognitive performance, and outlines phenomena including the irrelevant speech effect, as well as factors related to the acoustic characteristics of speech and their potential influence on performance and perceived workload. The aim is to provide a structured overview of these concepts and how they are interpreted within the context of this work.

2.1 Working Memory and Cognitive Performance

Cognition refers to the mental processes involved in interpreting, storing, and using information, which allows individuals to perceive, reason, and respond to their environment [12]. Processes within cognition include attention, memory, and executive control, which support the maintenance and manipulation of information during task performance.

Within this broader framework, working memory is commonly described as a system responsible for the temporary storage and manipulation of information. Although there is no single universally established definition [13], working memory is generally characterized by its dual role in maintaining information while simultaneously supporting ongoing cognitive processing.

Baddeley and Hitch [14] proposed the concept of working memory as an expansion of the traditional idea of short-term memory. Instead of a single temporary store, they argued for a multicomponent system that includes both short-term storage and a processing component, which they collectively referred to as working memory. This framework was elaborated by Baddeley [15], who introduced two passive storage systems: the phonological loop, responsible for maintaining speech-based information, and the visuospatial sketchpad, responsible for maintaining visual and spatial information, both of which are controlled and coordinated by the central executive. The episodic buffer provides temporary storage across the subsystems and long-term memory [16].

In contrast, Daneman and Carpenter [17] conceptualized working memory more as a single pool of resources that supports both storage and processing. Their work emphasized that, to measure working memory capacity, tasks must include a processing component so that the results are not solely based on short-term memory.

As noted above, there is still no complete consensus on how working memory should be defined, which underscores the need to interpret working memory measures in relation to the theoretical framework in which they are used. In this context, storage refers to the short-term maintenance of information, such as remembering a sequence of words, while processing involves parallel cognitive operations, such as solving mathematical equations or making decisions about presented stimuli. These two components compete for shared attentional resources, meaning that increased processing demands can reduce the ability to maintain stored information, and vice versa. Tasks such as the Operation Span Task (OSPAN) [10] are designed to capture this interaction by requiring participants to alternate between processing (e.g., solving mathematical operations) and storage (e.g., recalling memorized items), thereby assessing the capacity of the working memory. The task is described in detail in Section 3.3.1.

Maintaining information in working memory is not a passive process. It requires active attention control mechanisms to sustain and refresh the stored information over short periods, preventing them from being disrupted or displaced by new incoming information. This maintenance can be achieved through processes such as rehearsal or attentional focusing, which help preserve both the items that were presented and the order in which they occurred. However, because working memory capacity is limited, these processes are inherently vulnerable to interference. It has been argued that these limitations primarily reflect differences in attentional control, rather than storage capacity alone, emphasizing the role of maintaining goal-relevant information during distraction [18].

External stimuli, particularly those that are dynamic and attention-capturing, can disrupt the balance between storage and processing. Such interference may compete for attentional resources, introduce competing information, or interrupt the processes required to maintain the items and their order in memory. As a result, task performance can deteriorate even when the interfering stimulus is irrelevant to the task.

This vulnerability of working memory to external interference is central to understanding how auditory environments, and in particular speech exposure, can influence cognitive performance.

2.2 Speech Exposure and Cognitive Processing

In many everyday environments, individuals are regularly exposed to background sounds, such as noise from ventilation, traffic, and office equipment, among which speech is one of the most prominent. Unlike many other environmental sounds, speech varies continuously over time, reflecting the production of phonemes, syllables, and words.

In experimental and real-world task settings, speech is often present as task-irrelevant input, meaning that it does not contribute to the task being performed. At the same time, speech inherently contains structured information associated with communication. The processing of speech therefore involves the analysis of time-varying acoustic patterns in order to extract information from the signal. Contemporary accounts of speech perception describe this process as not purely passive, but influenced by attentional mechanisms and contextual factors [19].

As a result, the presence of speech might engage cognitive resources even when it is not relevant to the task. This creates conditions under which auditory inputs are processed despite being task-irrelevant, and may interact with task-related cognitive processing.

In the context of tasks that rely on working memory, this implies that speech exposure could engage perceptual and attentional resources required for maintaining and manipulating information. As discussed in the previous section, working memory relies on limited resources and is sensitive to interference from external input. Speech, as an information-bearing and continuously varying signal, therefore represents a potential source of such interference. The mechanisms by which speech influences cognitive performance, as well as the role of specific properties of the speech signal, are discussed in the following sections.

2.2.1 Irrelevant Speech Effect

When performing a cognitively demanding task while simultaneously being exposed to background speech, task performance is often disrupted and can occur even when the speech is entirely irrelevant to the task. This phenomenon, known as the irrelevant speech effect (ISE), refers to the finding that task-irrelevant speech interferes particularly with tasks that require the serial recall of items [6], [7].

Importantly, the disruptive effect of speech does not depend solely on its intelligibility or semantic content. Studies have shown that foreign-language speech, syntactically anomalous speech, and even sequences of non-words can impair performance to a similar extent as meaningful speech [7], [9], [20]. These findings indicate that the presence of structured, time-varying auditory input is sufficient to produce interference, suggesting that the effect is not exclusively driven by semantic processing.

The irrelevant speech effect is particularly relevant because many everyday tasks, such as reading, writing, or remembering instructions, depend on the ability to maintain and recall sequences of information in the correct order. The effect therefore provides a framework for understanding why such tasks are vulnerable to background speech and why certain types of sound are more disruptive than others.

From a cognitive perspective, maintaining sequences of items relies on working memory mechanisms that support both the temporary storage of information and the preservation of its order. When background speech is present, it may interfere with these processes in multiple ways. Within the multi-component working memory framework, visually presented items are often recoded into a speech-based format and maintained through subvocal rehearsal in the phonological loop. Background speech thereby introduces competing auditory information within this internal representation, disrupting the maintenance of the items and their order [7], [15].

In addition to such representational interference, the processing of dynamic auditory input may also engage the limited cognitive resources that support the parallel storage and processing of information. As a result, fewer resources remain available for maintaining task-relevant items, leading to performance decrements [17]. Notably, sounds that exhibit greater temporal variability, such as speech, tend to produce stronger disruption, highlighting the role of acoustic dynamics in the effect.

2.2.2 Acoustic Characteristics of Speech and Their Cognitive Effects

Speech is a complex acoustic signal composed of multiple interacting characteristics that influence how it is perceived and processed. These characteristics include temporal and spectral variability, intelligibility, and sound level, each representing a different aspect of the speech signal. In the context of cognitive performance, these characteristics are relevant because they influence how speech interacts with cognitive processes involved in maintaining and manipulating information. The following paragraphs describe three of such characteristics that are used to characterize speech in relation to cognitive performance and that form the basis for the experimental manipulations of the speech signals for this work.

Changing-state characteristics As discussed previously, the disruptive effect of task performance by background speech has been linked to the continuously varying structure of the speech signal. This property can be referred to as changing-state characteristics and describes the extent to which an acoustic signal changes over time.

Speech consists of sequences of phonemes, syllables, and words, each introducing new acoustic information. These ongoing changes result in a stream of discrete auditory events that can interfere with the processes involved in maintaining ordered information in working memory. In tasks that rely on serial recall, where items must be retained in a specific sequence, such variation introduces additional elements that compete with the internally maintained sequence.

This interpretation is supported by findings showing that disruption increases when the auditory signal contains a sequence of changing elements, compared to acoustically stable sounds [9], [21]. The effect is therefore not dependent on the presence of linguistic meaning, but on the extent to which the signal introduces successive changes over time. Each change in the signal can be understood as adding new, task-irrelevant information that must be processed or suppressed, thereby interfering with the maintenance of task-relevant items.

Consistent with this, signals that lack such variation, such as steady-state noise, typically produce less disturbance, indicating that it is the temporal structure of the signal rather than its mere presence that drives the effect [9], [20], [22]. Figure 2.1 shows a simple illustration of the difference in changing-state characteristics between speech and a steady-state noise signal.

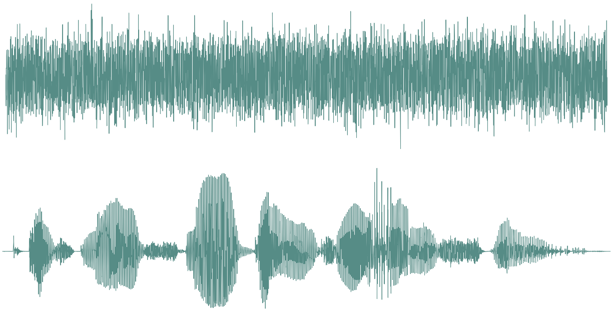

Figure 2.1: Conceptual illustration of the difference in changing-state characteristics of a white noise signal (top) and a speech signal (bottom).

Intelligibility In addition to the temporal variation described above, the extent to which a speech signal can be understood also influences how it interacts with cognitive processing. Speech intelligibility refers to the degree to which the linguistic content of the signal can be perceived and interpreted. It is influenced not only by the speech signal itself, but also by acoustic conditions within the environment, such as background noise and reverberation, which can reduce the clarity and temporal definition of the speech signal.

Although changing-state characteristics describe how variation in the signal over time interferes with the maintenance of ordered information, intelligibility relates to the degree to which the signal can be processed at a linguistic level. Differences in intelligibility have been associated with differences in cognitive performance, with more intelligible speech generally producing stronger interference in tasks that involve working memory [22], [23], [24]. This relationship could be attributed to the increased likelihood that intelligible speech engages linguistic processing, thereby increasing its interaction with task-related cognitive processes.

As discussed in the previous sections, disruption can occur even when the speech signal is not meaningful, indicating that intelligibility is not a necessary condition for interference. Rather, intelligibility represents an additional dimension that can influence the magnitude of the effect when present.

Sound level Sound level describes the amplitude of the speech signal and determines its audibility within the listening environment. It defines the conditions under which the signal is present relative to the ongoing task.

Changes in sound level affect how easily the speech signal can be perceived and, consequently, whether the variations and structure described in the preceding sections are available to the listener. At lower levels, parts of the signal might fall below perceptual relevance, whereas higher levels increase the probability that elements of the speech signal are detected and followed over time.

Differences in sound level have been examined in relation to cognitive performance, but the evidence does not show a consistent relationship. Several studies report limited or negligible effects within certain level ranges tested [21], [23]. This suggests that any influence of sound level on cognition is likely to arise through its modulation of how perceptually available the acoustic properties of speech are, rather than through a distinct or independent form of interference.

2.3 Listening Effort and Perceived Workload

So far, the effects of background speech have been described in terms of how it influences cognitive processing, particularly working memory, and the mechanisms involved in maintaining and manipulating information. These effects are often reflected in objective measures such as recall performance. However, such measures do not fully capture how demanding a task is experienced.

Exposure to background speech has been shown to increase perceived workload and listening effort, even when the speech is not relevant to the task being performed [5], [23]. Subjective evaluations in these conditions often indicate mental demand and effort, suggesting that the presence of speech alters the experienced difficulty of the task, independent of observable changes in performance.

A consistent observation is that subjective workload and objective performance do not necessarily vary in parallel [25]. Stable performance can be maintained under speech exposure, while subjective ratings indicate increased effort or strain [23]. This suggests that individuals may adjust their cognitive processing in order to preserve task performance in the presence of distracting input.

This pattern can be interpreted within capacity-limited models of attention, where cognitive processing is constrained by a finite pool of resources that can be flexibly allocated depending on task demands [26]. When background speech is present, additional resources may be required to maintain task-relevant processing and to manage interfering input. Similarly, effortful listening frameworks describe listening as an active process in which cognitive resources are allocated to maintain performance under challenging conditions [27]. In this context, increased listening effort reflects the additional processing required to extract or suppress information from the auditory signal.

The maintenance of stable performance under increased demand is often attributed to compensatory mechanisms, where additional cognitive resources are recruited to counteract the effects of distraction [28]. While such compensation can preserve performance in the short term, it is associated with an increased subjective cost, reflected in a higher perceived workload. Over time, sustained reliance on these mechanisms is likely to lead to fatigue and increased strain.

Together, these findings indicate that subjective measures provide complementary information to objective performance measures. They capture aspects of cognitive demand that are not directly reflected in task outcomes and are therefore necessary for a more complete understanding of how speech exposure affects individuals during task performance [25].

Chapter 3

Acoustic and Experimental Framework

When examining phenomena related to auditory perception and cognitive performance, such as the effects of speech exposure, it is necessary to represent the acoustic conditions under which such exposure occurs. In this work, controlled experimental conditions are used to approximate relevant aspects of real-world environments while allowing systematic manipulation of specific variables. This chapter describes the acoustic and experimental framework underlying the work presented in this thesis. It outlines how acoustic environments were simulated, how speech stimuli were generated and manipulated, and how the listening experiments were designed and conducted. In addition, the cognitive and subjective measures used to evaluate performance and perceived workload are presented.

3.1 Simulation of Acoustic Environments

In this work, acoustic environments are simulated to approximate key characteristics of real spaces while enabling systematic variation of specific parameters. This section describes how the acoustic environments are modeled, including both room-related and listener-specific acoustic properties.

3.1.1 Room Acoustic Properties

All acoustic properties of an enclosed space can, in principle, be described by its room impulse response (RIR), which characterizes how the environment modifies an incoming sound signal as it propagates from a source to a receiver. The room impulse response provides a linear description of the acoustic system and contains information about the temporal and spectral modifications of the sound-field, including energy decay characteristics, frequency dependency, and the temporal distribution of reflections, from which parameters such as reverberation time can be derived.

In a conceptual sense, the room impulse response can be decomposed into three primary components: the direct sound, the early reflections, and the late reverberant field. The direct sound corresponds to the sound energy that travels along the shortest path from the source to the listener without interacting with the surrounding environment. Following the direct sound, early reflections arise from first- and higher-order reflections from nearby surfaces such as walls, floor, and ceiling.

As the number of reflections increases, the sound field becomes progressively more diffuse, resulting in a dense collection of reflected sound energy commonly referred to as the reverberant tail. At this point, individual reflections are no longer distinguishable, and the sound energy decays over time. The characteristics of both the early reflections and the late reverberant field are determined by the geometry of the space, the relative positions of the source and receiver, and the absorption and scattering properties of the enclosing surfaces. Consequently, parameters such as reverberation time, typically defined as the time required for the sound pressure level to decay by 60 dB after the source has stopped, and temporal distribution of reflections are directly linked to the physical properties of the room.

Room impulse responses can be obtained either through direct measurement or through numerical simulation. In measurement-based approaches, excitation signals such as impulses or sine sweeps are emitted into the space, and the resulting response is recorded at the receiver position. Alternatively, computational models can be used to generate room impulse responses based on geometrical and material properties of the environment, see Section 3.1.1.1 for a description of the method used for the acoustic conditions in this thesis. In both cases, the resulting room impulse response can be used to reproduce the acoustic characteristics of the space, for example, by combining it with audio signals through convolution for auralization purposes.

3.1.1.1 Modeling Room Impulse Responses

The room impulse responses used in the experiments in the appended papers were generated using a hybrid prediction tool combining an image source model with an acoustic radiosity model. The approach is more thoroughly explained in [29].

The image source model describes sound propagation by modeling reflections at room boundaries as contributions from a set of virtual sources, referred to as image sources. These sources are constructed by mirroring the original source position across the room boundaries, such that each image source represents a specific reflection path between the source and the receiver. Higher order image sources therefore represent increasingly complex propagation paths.

The total number of sources, including the real source, is determined by

$$N_s = (2N_{\text{ord}} + 1)^3, \tag{3.1}$$

where the image sources are indexed by n_x, n_y, and n_z according to

$$-N_{\text{ord}} \leq n_x, n_y, n_z \leq N_{\text{ord}} \tag{3.2}$$

which is illustrated conceptually in Figure 3.1. For each image source, the contribution to the receiver is determined by the geometric distance between the source and the receiver, which defines both the propagation delay and the amplitude decay. In addition, each reflection along the propagation path introduces further attenuation according to the reflection properties of the corresponding boundary surfaces. The total sound pressure at the receiver is obtained by summing all such contributions, resulting in a superposition of the direct sound and a set of reflections.

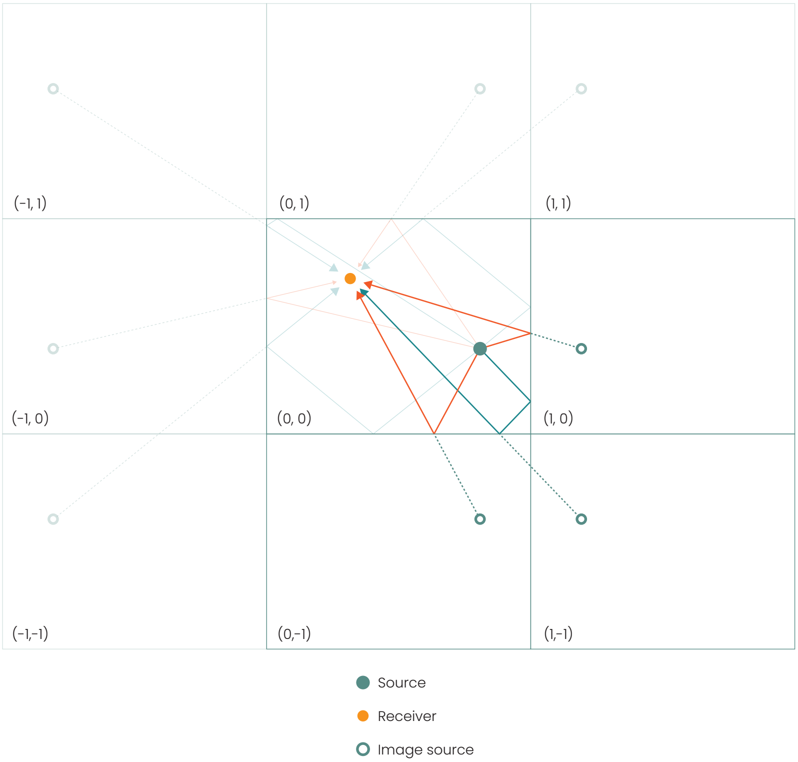

Figure 3.1: Simple 2D illustration of the image source model geometry, where the indices denote n_x and n_y.

This model provides an accurate description of the direct sound and early reflections, where individual propagation paths can be resolved in time. However, as the reflection order increases, the number of image sources grows rapidly, making the method computationally heavy and less suitable for modeling the late reverberant sound field.

To address this, the acoustic radiosity model is used to describe the late, diffuse part of the sound field. In contrast to the image source model, the radiosity model does not track sound as waves but instead represents it in terms of energy. Room boundaries are discretized into small surface elements, each of which receives incoming sound energy, absorbs a portion of it, and reflects the remaining energy diffusely into the room. The outgoing energy from each surface element is described by the radiation density B_i, which represents the amount of sound energy leaving element i. By solving the resulting system of energy balance equations, the model determines how sound energy is redistributed throughout the room, allowing the diffuse sound field and late reverberation to be estimated.

The image source model preserves phase information and allows interference effects between different propagation paths, which is important for accurately modeling the direct sound and early reflections. In contrast, the radiosity model is based on energy quantities and does not retain phase information. For the late, diffuse sound field, however, phase relations between contributions are effectively random, making an energy-based description sufficient and computationally more efficient than a wave-based approach.

In the hybrid method [29], the two models are coupled through a reflection-based energy partitioning. At each reflection of the image source model, the incident sound is divided into three components: an absorbed part, a reflected part, and a scattered part. The absorbed fraction is determined by the absorption coefficient α, while the scattered fraction is determined by the scattering coefficient σ. The remaining component, given by $(1-\alpha)(1-\sigma)$, continues to propagate within the image source model as a reflection. In contrast, the scattered component, $(1-\alpha)\sigma$, is removed from the image source model and transferred to the radiosity model, where it is treated as diffuse energy. Once transferred, this energy remains within the radiosity framework and is redistributed between surfaces according to the radiosity formulation. The distribution and decay of this diffuse energy field are governed by the absorption properties of the room surfaces, and therefore determine the overall energy decay of the impulse response. Consequently, the reverberation time is controlled through the choice of absorption and scattering coefficients. For a given source-receiver configuration, the simulation produces 72 directional impulse responses at 5° azimuthal resolution, corresponding to a horizontal-plane representation of the sound field.

3.1.2 Listener-specific Acoustic Properties

In addition to the room-specific acoustic properties, the characteristics of the listener also play a crucial role in determining how sound is perceived. As sound propagates from a source to a listener, it is not only reflected by surfaces in the environment, but also interacts with the listeners torso, head, and ears. These interactions modify the incoming sound before it reaches the ear canals.

Furthermore, due to the spatial separation between the ears, the sound arrives at the left and right ears with small differences in both time and level. These interaural time differences (ITDs) and interaural level differences (ILDs), together with spectral modifications introduced by the outer ear, provide important cues for spatial hearing. By comparing the signals arriving at each ear, the auditory system is able to interpret the direction and, to some extent, the distance of a sound source.

The transformation imposed by the listeners anatomy can be described by the head-related transfer function (HRTF), which characterizes how an incoming sound from a specific direction is filtered by the torso, head, and pinnae before reaching the eardrum. The HRTF is inherently individual, as it depends on the geometry and dimensions of the listeners body and ears, leading to variations in how sound is perceived between individuals. Since individual HRTFs are rarely available, standardized databases of averaged or representative HRTFs are used to approximate a typical listener.

3.1.3 Binaural Room Impulse Response

In practice, listener-specific acoustic acoustic transformations are incorporated into simulated environments by combining room impulse responses with head-related transfer functions. This is achieved by convolving the room impulse response with a pair of HRTFs corresponding to the left and right ear for a given direction of arrival, resulting in a two-channel binaural room impulse response (BRIR). The BRIR thus represents the signals that would be received at the two ears for a specific source-receiver configuration within the simulated environment.

In the present work, the directional impulse responses generated by the hybrid prediction tool are transformed into binaural representations suitable for headphone-based reproduction. Each of the 72 directional impulse responses is convolved with the corresponding HRTF [30] associated with its direction of arrival. The resulting BRIRs are combined and subsequently used to render the experimental stimuli, allowing controlled acoustic conditions to be reproduced while preserving spatial characteristics of the sound field.

3.2 Speech Stimuli and Acoustic Manipulations

Speech Stimuli The speech stimuli used in this thesis were constructed from short utterances, with an average length of 2.4 s, obtained from the McGill University TSP Speech Database [31]. The recordings consist of English sentences spoken by both male and female speakers under anechoic conditions, ensuring that the signals are free from room acoustic effects prior to further processing.

The text material for the utterances is obtained from the Harvard sentences [32]. The sentences are designed to provide a representative sample of speech sounds without introducing strong lexical or semantic predictability, which makes them suitable for experiments where consistent and controlled speech material is required across conditions. Examples of such sentences include

THE BIRCH CANOE SLID ON THE SMOOTH PLANKS

THE BOY WAS THERE WHEN THE SUN ROSE

THE SMALL PUP GNAWED A HOLE IN THE SOCK

By using these utterances, it is possible to generate speech stimuli that are acoustically representative of natural speech while maintaining a high degree of experimental control. In particular, the use of phonetically balanced and semantically neutral sentences reduces variability related to linguistic content, allowing the effects of acoustic manipulations to be examined more systematically.

The speech stimuli in the listening experiments in this thesis were constructed by processing and concatenating the speech utterances to form continuous signals. To ensure consistency across stimuli, each speech snippet was first normalized to a common sound level. This normalization procedure removes level variations between individual recordings and speakers, ensuring that the differences in the stimuli are not driven by unintended amplitude variations but instead reflect the intended experimental manipulations. To avoid artifacts in the transitions between segments, short fade-in and fade-out windows were applied to each snippet.

The order of the speech snippets was randomized for each stimuli and concatenated to form longer sequences. A pause of 0.5 seconds was added between consecutive snippets, which preserves the integrity of each sentence, while also introducing temporal separation that reflects natural conversational patterns. To maintain variability while avoiding long sequences on a single voice, snippets from different male and female speakers were alternated throughout the stimuli. This provides a more dynamic speech signal, reducing potential adaptation effects and limiting speaker-biases.

Locally Time-Reversed Speech To examine the role of semantic content in speech-related distraction, locally time-reversed speech stimuli were generated. This manipulation removes linguistic intelligibility while preserving short-term acoustic cues and characteristics of speech, such as temporal fluctuations and spectral structure of the original speech signal. This allows the contribution of semantic information to be isolated from acoustic properties. The locally time-reversed stimuli were constructed by segmenting the original speech signals into short temporal windows, reversing each segment in time and subsequently concatenating them in their original sequence. Prior to concatenation, each segment was windowed to reduce discontinuities and avoid the introduction of artifacts at segment boundaries.

Segment durations were randomly varied between 100 and 300 ms. The selection of segment lengths was based on findings showing that speech intelligibility deteriorates when temporal segments are reversed beyond approximately 100 ms, while shorter segments may still preserve partial intelligibility [33]. By using segments in the range of 100-300 ms, the resulting stimuli are rendered largely unintelligible while maintaining local temporal and spectral characteristics of speech. The use of randomized segment lengths further reduced the potential for periodicity or predictability in the signal, ensuring that the manipulations primarily affect intelligibility rather than introducing additional structured artifacts.

A simple perceptual illustration of the segmentation and time-reversal procedure is shown in Figure 3.2.

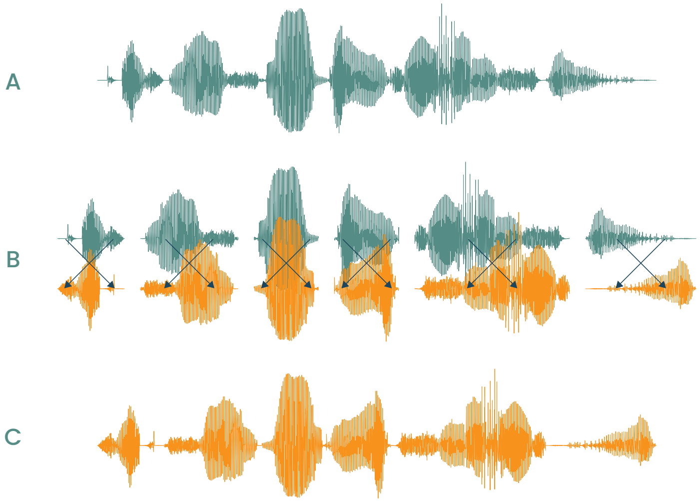

Figure 3.2: Conceptual illustration of how the locally time-reversed speech stimuli were created. A represents the original speech stimuli. B represents the segmentation and time-reversal of each segment. C represents the final concatenation of the time-reversed segments.

Sound Level In order to examine the effects of the sound level of speech on cognitive performance and subjective workload, the original speech stimuli were processed to obtain different presentation levels. This was achieved by applying a uniform amplitude scaling to the signals, adjusting their root-mean-square (RMS) level to predefined target values. This modifies the overall sound pressure level of the stimuli, ensuring that any observed effects can be attributed to level differences rather than changes in signal structure. The applied levels were selected to represent a range of conditions relevant to typical acoustic environments, from low-level background speech to more prominent speech signals. The resulting levels were verified using binaural recordings and analyzed in terms of A-weighted sound pressure level.

3.3 Cognitive and Subjective Measures

To evaluate the effects of speech exposure on both performance-related effects and experienced effort under different acoustic conditions, an approach with objective and subjective measures was employed. Cognitive performance was assessed using a cognitive task designed to capture both storage and processing demands, while perceived workload was assessed using a multi-dimensional self-report measure.

3.3.1 Operation Span Task

The Operation Span Task (OSPAN) [10] is a well-established measure of working memory, designed to assess an individual's ability to simultaneously store and process information. In the task, participants are required to remember a sequence of items and subsequently recall them in the correct order. Tasks that combine storage and processing, such as the Operation Span Task, are particularly sensitive to auditory distraction. Because OSPAN requires participants to maintain verbal items while simultaneously performing processing operations, any additional demand, such as processing background speech, further taxes the limited cognitive resources available.

In the experiments in the present thesis, the to-be-remembered items consisted of words, presented in random order, containing 2-3 syllables and 5-7 letters, as per examples below.

HONEY
OPTIMAL
EASILY
RADISH
JUROR
TWINKLE
VERIFY
MAFIA
INVADER
RETIRED
FINGER
POTATO

These were presented sequentially and interleaved with processing tasks in the form of mathematical operations, where participants were required to determine whether a given equation was true or false. The operations serve to disrupt the continuous rehearsal of the memory items, thereby increasing the cognitive demands of the task. The equations were generated using a randomization procedure following the structural form

$$N_1 \times N_2 \pm N_3 = R, \tag{3.3}$$

where $N_1, N_2, N_3 \in \{1, 2, ..., 10\}$ and with addition and subtraction assigned in equal proportion. Half of the equations were designated true, in which case the result R was computed exactly. The remaining half were false, where R was obtained by adding a random integer offset between -3 and $+3$ to the true result, excluding zero to ensure incorrectness.

Each trial consists of a series of equation-word pairs, after which participants are prompted to recall the words. The number of pairs within a trial varies, resulting in varying memory loads across trials. This structure ensures that the task captures both storage and processing components of working memory, as participants must maintain the sequence of words while concurrently solving the intervening operations. For the experiments in this thesis, equation-word pairs ranged from three to eight.

Performance in the OSPAN task is commonly quantified based on the number of correctly recalled items, using scoring approaches based either on strict serial recall or recall independent of order.

A conceptual illustration of the OSPAN procedure is presented in Figure 3.3 and graphical examples of the implementation are provided in Paper II.

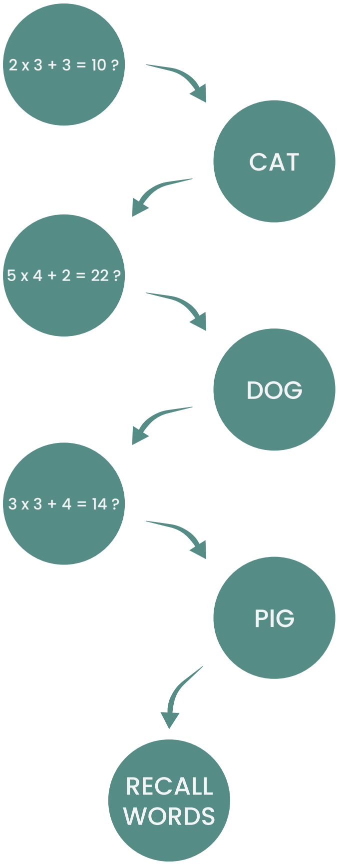

Figure 3.3: Conceptual illustration of the Operation Span Task (OSPAN) procedure.

3.3.2 NASA Task Load Index

The NASA Task Load Index (NASA-TLX) [11] is a widely used subjective assessment tool for evaluating perceived workload during task performance. It provides a multi-dimensional measure of workload by capturing participants self-reported experiences across several contributing factors. In experiments involving multiple conditions, the questionnaire is typically administered after each condition to allow for comparative evaluation of perceived task demands.

The NASA-TLX consists of six sub-scales representing different dimensions of workload

MENTAL DEMAND	How mentally demanding was the task?
PHYSICAL DEMAND	How physically demanding was the task?
TEMPORAL DEMAND	How hurried or rushed was the task?
PERFORMANCE	How successful were you in accomplishing what you were asked to do?
EFFORT	How hard did you have to work to accomplish your level of performance?
FRUSTRATION	How insecure, discouraged, irritated, stressed, and annoyed were you?

Each sub-scale is rated on a continuous scale ranging from *Very Low* to *Very High*, except for the performance scale, which is rated between *Perfect* and *Failure*. Although the original implementation employed a 21-point scale derived from a subdivided Likert-type format, responses are commonly treated as continuous and mapped onto a range from 0 to 100, where higher values indicate greater perceived workload.

To obtain an overall workload score, the NASA-TLX incorporates a weighting procedure that reflects the relative contribution of each sub-scale to the perceived workload. This is achieved through pairwise comparisons between all sub-scales, resulting in 15 comparisons. In each comparison, participants indicate which of the two dimensions contributed more to their perceived workload. The number of times a sub-scale is selected determines its weight, with a maximum possible weight of five.

The final workload score is calculated by multiplying each sub-scale rating by its corresponding weight and summing the weighted values. The resulting score is then normalized to a scale from 0 to 100, where 0 represents minimal workload and 100 represents maximal workload. In practice, the weighting procedure is often performed once per participant after completing all experimental conditions, as it is assumed to reflect stable task-related priorities [34].

This multi-dimensional approach allows the NASA-TLX to capture both the overall level of perceived workload and the relative importance of different contributing factors. Graphical examples of the implementation of the NASA-TLX are provided in the Paper II.

3.4 Experimental Paradigm

The listening experiments conducted in this thesis employed a within-subjects design, in which each participant was exposed to all experimental conditions within each listening experiment. This approach allows for direct comparison of performance across conditions while controlling for inter-individual variability. The experimental procedure was identical across all experiments, with the exception of the acoustic conditions to which participants were exposed.

All experiments were implemented in a custom-developed program using MATLAB, which controlled stimulus presentation, task execution, and data collection. Prior to the start of the experiment, participants were informed about the procedure and the tasks involved. Informed consent was obtained from all participants and they were instructed that participation was voluntary and could be terminated at any time.

Participants first completed a brief registration phase, in which basic demographic information was collected, including participation code (for anonymization), age, gender, and self-reported hearing loss. Participants who reported hearing impairment were excluded from the results. Before the main experiment, participants completed a short practice session of the Operation Span Task (OSPAN) to familiarize themselves with the task structure and interface. This session also provided an opportunity to ask questions and ensure that the instructions were clearly understood.

During the experiment, each participant completed four rounds of the OSPAN task. Within each round, the sequence lengths of the equation-word pairs ranged from three to eight, yielding 24 sets in total. Each round was performed under a specific acoustic condition, including both active sound conditions and a silent baseline condition. The order of presentation was randomized across participants to mitigate potential order effects and to prevent that time-dependent influences are systematically associated with a specific condition.

Immediately following each OSPAN round, participants completed the NASA Task Load Index (NASA-TLX) questionnaire, reporting their perceived workload for the condition they had just experienced. This allowed subjective workload to be assessed in direct relation to each experimental condition. After all rounds had been completed, participants performed the pairwise comparison procedure for the NASA-TLX weighting, which determined the relative contribution of each workload dimension to the overall score.

3.4.1 Acoustic Conditions

Each listening experiment comprised four acoustic conditions, including a silent baseline condition used as a reference for performance and perceived workload in the absence of auditory stimulation.

The first listening experiment, presented in Paper I, examined the influence of reverberation time. Two reverberation conditions were included, representing a short (0.2 s) and a long (1.0 s) reverberation time, both combined with speech stimuli. In addition, one condition included speech with short reverberation time combined with typical office-related background sounds. The resulting conditions therefore consisted of three sound conditions (two with short reverberation time, one with long reverberation time) and one silent baseline.

The second listening experiment (Listening Experiment A in Paper II examined the role of speech characteristics. The conditions included intelligible speech, locally time-reversed speech, Brownian noise, and a silent baseline condition.

The third and fourth listening experiments (Listening Experiments B1 and B2 in Paper II) examined the influence of sound level. In both experiments, three speech conditions with different presentation levels were used together with a silent baseline condition. In B1, the levels spanned a range of 31 dB(A) to 61 dB(A), while B2 focused on a more narrow range of 31 dB(A) to 43 dB(A).

Chapter 4

Overview of the Appended Papers

This chapter provides summaries of the appended papers included in the thesis. The papers investigate how different room-acoustic and speech-related variables influence working memory performance and perceived workload under controlled listening conditions, using the Operation Span Task (OSPAN) and NASA Task Load Index (NASA-TLX) within a common experimental and methodological framework as explained in previous chapters.

For each paper, a brief overview of the methodology, results, and discussion is presented, while the complete studies are included in Part II. A broader discussion of the overarching findings and their implications is presented in Chapter 5.

4.1 Influence of Speech Exposure on Working Memory: Exploring Reverberation Time in Open Plan Office Environments

In this paper, published as a conference paper in *Fortschritte der Akustik - DAGA 2024: 50 Jahrestagung für Akustik*, results from a laboratory listening experiment are presented, examining the influence of reverberation time on working memory and perceived workload during speech exposure.

The cognitive effects of speech exposure have been widely studied, and several factors have been identified as influencing how speech affects cognitive performance. In realistic environments, speech is rarely presented under anechoic conditions, and room acoustics, including reverberation time, can alter both the intelligibility and temporal characteristics of speech. Since more intelligible speech has often been associated with stronger cognitive disruption, reverberation could indirectly influence how strongly background speech interferes with working memory performance. These aspects are therefore relevant to consider when examining how speech exposure affects cognitive performance.

This paper investigates the role of reverberation time in shaping the cognitive effects of speech exposure. By systematically varying reverberation under controlled conditions, the study provides insight into how room-acoustic properties influence both working memory performance and perceived workload. The acoustic environments were simulated using binaural room impulse responses (BRIR) with reverberation times of 0.2 s and 1.0 s.

The results showed that exposure to speech reduced working memory performance and increased perceived workload compared to silence. However, no significant differences were observed between the reverberation time conditions in either the objective or subjective measures. These findings suggest that the disruptive effect of speech may be more strongly related to the properties of the speech signal itself than to the reverberation times examined in the study.

4.2 Influence of Speech Characteristics on Working Memory and Perceived Workload: Findings from Experimental Listening Experiments

In this paper, currently under review, results from three laboratory listening experiments are presented, examining how semantic content and sound level influence working memory and perceived workload during speech exposure.

Although the disruptive effects of background speech are well established, it remains unclear which and to what extent specific properties of speech drive these effects. In particular, the relative contributions of semantic content, acoustic structure, and sound level are not fully disentangled. This makes it difficult to determine whether cognitive disruption is primarily driven by linguistic processing or by more general acoustic characteristics of speech.

This paper investigates how different characteristics of speech, including semantic content and sound level, influence working memory performance and perceived workload. By isolating these factors across multiple experiments, the study contributes to a more detailed understanding of which aspects of speech are most relevant for cognitive disruption. The study consisted of three laboratory listening experiments in which semantic content and sound level were systematically varied using speech stimuli presented under controlled acoustic conditions.

Across the experiments, speech exposure was associated with reduced working memory performance and increased perceived workload relative to silence. Higher sound levels were associated with stronger disruption, while speech-like signals lacking semantic content produced effects comparable to intelligible speech. These findings suggest that the disruptive effect of speech is not heavily determined by semantic content but is more strongly influenced by acoustic characteristics inherent to speech signals as well as its perceptual prominence.

Chapter 5

Conclusions

The implications of speech exposure for working memory and perceived workload involve several interacting perspectives, including cognitive psychology, auditory perception, room acoustics, and occupational acoustics. Within these areas, speech-related distraction can be interpreted through different theoretical frameworks, ranging from working-memory models and the irrelevant speech effect to acoustic descriptions of speech intelligibility, temporal variability, and sound level. The papers included in this thesis contribute to this broader field by examining speech exposure not only as a general disturbance, but as a phenomenon shaped by both acoustic conditions and cognitive responses.

Taken together, the results show that background speech affects both subjective and performance-based aspects of cognition. Across the included studies, speech exposure was associated with reduced working memory performance and increased perceived workload, indicating that speech imposes a cognitive cost under controlled laboratory conditions. However, the results also show that this cost is not determined by speech exposure alone. Instead, the effect depends on how speech is acoustically represented, including its level, speech-like temporal structure, and the listening conditions in which it occurs. In contrast, the reverberation time conditions examined in Paper I did not produce clear differences in the experimental measures, suggesting that the influence of reverberation time might be less pronounced under the specific listening conditions considered here. This could indicate that modifying room-acoustic conditions alone is not necessarily sufficient to reduce cognitive disruption if speech-related characteristics such as speech prominence or temporal variability remain largely preserved.

The overarching contribution of the thesis is not the demonstration that speech can be disruptive, which is already well established, but the systematic examination of how different acoustic characteristics shape this disruption and how objective and subjective outcomes relate to one another. The work shows that speech-like structure can be disruptive even when semantic content is reduced, while acoustically stable background sounds produced smaller effects.

Together, these findings suggest that the disruptive effect cannot be understood solely in terms of semantic processing or the mere presence of sound. Instead, characteristics associated with temporal variation and speech-like structure appear to contribute substantially to the effect, likely because continuously changing auditory signals repeatedly engage cognitive and attentional processes during task performance. The results further show that sound level modulates the strength of the effect and that subjective workload may increase even when performance effects are small or absent. This contributes to a more nuanced interpretation of speech exposure, where cognitive disruption is understood as the result of an interaction between acoustic properties, task demands, and compensatory effort.

These findings are consistent with theoretical accounts in which working memory is vulnerable to interference from task-irrelevant auditory input, particularly when the sound contains temporal variation. They also align with capacity-based accounts of attention, where individuals can maintain performance by allocating additional cognitive resources. From this perspective, stable performance should not necessarily be interpreted as absence of disruption, since the cost of maintaining performance may instead appear in subjective workload. The observed variability between participants also indicates that susceptibility to speech-related distraction can differ substantially between individuals and that the relationship between acoustic conditions and cognitive disruption is unlikely to be uniform across listeners.

At the same time, the conclusions should be interpreted within the scope of controlled listening experiments. The acoustic environments were simulated, the tasks were short and well defined, and the participants were exposed to speech through headphones rather than in fully naturalistic office settings. Although the findings remain relevant for environments in which sustained cognitive work occurs in the presence of background speech, translating laboratory findings to real-world environments require consideration of the complexity of everyday listening situations. Real work environments include additional factors, such as visual context, social meaning, multiple simultaneous sound sources, task variability, and longer exposure durations, which could all influence how speech is perceived and attended to, as well as affect the cognitive resources required to maintain task performance under speech exposure.

Chapter 6

Future Work

Although the present work provides further insight into how acoustic characteristics of speech influence cognitive performance and perceived workload, several questions remain regarding how these findings generalize across listeners, tasks, and listening environments. Future work is therefore needed to better understand both the variability in individual responses to speech exposure and the conditions under which laboratory findings can be considered representative of real-world situations.

Regarding inter-individual variability in susceptibility to speech-related distraction, factors such as working memory capacity, attentional control, listening strategies, or prior exposure to noisy environments could influence how strongly speech affects both cognitive performance and perceived workload. An improved understanding of such individual differences may help explain why some listeners are more vulnerable to background speech than others under comparable acoustic conditions.

For laboratory findings to be transferable to real-world environments, future research must further examine under which conditions controlled listening experiments can be considered representative of real-world situations. This involves identifying which aspects of the acoustic environment, task design, and exposure conditions must be preserved for experimental findings to remain applicable outside the laboratory.

For example, it remains necessary to determine how factors such as the number of concurrent sound sources, spatial distribution of speech, exposure duration, and task characteristics influence the transferability of results. In addition, the extent to which simplified laboratory stimuli capture the perceptual and cognitive demands of real speech environments needs to be evaluated.

Rather than increasing realism in a general sense, this would require systematic investigation of which parameters are critical for maintaining ecological validity. Establishing such criteria would allow controlled experiments to be interpreted within a clearer framework, where the relationship between laboratory conditions and real-world environments is explicitly defined. This is essential for ensuring that conclusions drawn from experimental studies can be reliably applied to the design and assessment of acoustic environments in practice.

Bibliography

[1] W. Kropp, "Perceptionsbaserad planering av de akustiska förhållandena i kontorslandskap med hjälp av en virtuell miljö," Department of Architecture and Civil Engineering, Chalmers University of Technology, Gothenburg, Sweden, Tech. Rep. 190278, 2024, Accessed: 2026-04-09. [Online]. Available: `https://fou.afaforsakring.se/sv/forskningsarkivet/projekt/190278` (cit. on p. v).

[2] Arbetsmiljöverket, "Arbetsmiljön 2024," Arbetsmiljöverket, Stockholm, ISSN 1652-1110, Product No. AM0501 2025:01, 2025, Accessed: 2026-05-04. [Online]. Available: `https://www.av.se/globalassets/filer/statistik/arbetsmiljon-2024/arbetsmiljon-2024-rapport.pdf?hl=arbetsmilj%C3%B6n%202024` (cit. on p. 3).

[3] S. P. Banbury and D. C. Berry, Office noise and employee concentration: identifying causes of disruption and potential improvements, *Ergonomics*, vol. 48, pp. 25–37, 1 2005. DOI: `10.1080/00140130412331311390` (cit. on p. 4).

[4] A. Haapakangas, V. Hongisto, M. Eerola and T. Kuusisto, Distraction distance and perceived disturbance by noise - An analysis of 21 open-plan offices, *The Journal of the Acoustical Society of America*, vol. 141, pp. 127–136, 2017. DOI: `10.1121/1.4973690` (cit. on p. 4).

[5] H. Jahncke, S. Hygge, N. Halin, A. M. Green and K. Dimberg, Open-plan office noise: Cogntive performance and restoration, *Journal of Environmental Psychology*, vol. 31, pp. 373–382, 2011. DOI: `10.1016/j.jenvp.2011.07.002` (cit. on pp. 4, 14).

[6] H. A. Colle and A. Welsh, Acoustic Masking in Primary Memory, *Journal of Verbal Learning and Verbal Behavior*, vol. 15, pp. 17–31, 1976. DOI: `10.1016/S0022-5371(76)90003-7` (cit. on pp. 4, 11).

[7] P. Salamé and A. Baddeley, Disruption of short-term memory by unattended speech: Implications for the structure of working memory, *Journal of Verbal Learning and Verbal Behavior*, vol. 21, no. 2, pp. 150–164, 1982. DOI: `10.1016/S0022-5371(82)90521-7` (cit. on pp. 4, 11, 12).

[8] D. M. Jones and W. J. Macken, Irrelevant speech and serial recall, *Quarterly Journal of Experimental Psychology*, vol. 44, pp. 673–688, 1992. DOI: `10.1111/j.1467-9450.1992.tb00911.x` (cit. on p. 4).

[9] D. Jones, C. Madden and C. Miles, Privileged Access by Irrelevant Speech to Short-term Memory: The Role of Changing State, *Quarterly Journal of Experimental Psychology*, vol. 44A, pp. 645–669, 4 1992. DOI: `10.1080/14640749208401304` (cit. on pp. 4, 12, 13).

[10] M. L. Turner and R. W. Engle, Is Working Memory Capacity Task Dependent? *Journal of Memory and Language*, vol. 28, pp. 127–154, 1989. DOI: `10.1016/0749-596X(89)90040-5` (cit. on pp. 6, 10, 24).

[11] S. G. Hart and L. E. Staveland, Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research, in *Human Mental Workload*, ser. Advances in Psychology, vol. 52, North-Holland, 1988, pp. 139–183. DOI: `10.1016/S0166-4115(08)62386-9` (cit. on pp. 6, 26).

[12] M. W. Eysenck and M. T. Keane, Cognitive psychology: A student's handbook, 7th ed. Psychology Press, 2017 (cit. on p. 9).

[13] N. Cowan, The many faces of working memory and short-term storage, *Psychonomic Bulletin & Review*, vol. 24, 4 2017. DOI: `10.3758/s13423-016-1191-6` (cit. on p. 9).

[14] A. D. Baddeley and G. Hitch, Working Memory, in ser. Psychology of Learning and Motivation, G. H. Bower, Ed., vol. 8, Academic Press, 1974, pp. 47–89. DOI: `10.1016/S0079-7421(08)60452-1` (cit. on p. 10).

[15] A. D. Baddeley, Working Memory (Oxford Psychology Series 11). Oxford: Clarendon Press, 1986, ISBN: 9780198521334 (cit. on pp. 10, 12).

[16] A. Baddeley, The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, vol. 4, pp. 417–423, 11 2000. DOI: `10.1016/s1364-6613(00)01538-2` (cit. on p. 10).

[17] M. Daneman and P. A. Carpenter, Individual Differences in Working Memory and Reading, *Journal of verbal learning and verbal behavior*, vol. 19, 1980. DOI: `10.1016/S0022-5371(80)90312-6` (cit. on pp. 10, 12).

[18] R. W. Engle, M. J. Kane and S. W. Tuholski, Individual Differences in Working Memory Capacity and What They Tell Us About Controlled Attention, General Fluid Intelligence, and Functions of the Prefrontal Cortex, in cambridge University Press, 1999, pp. 102–134, ISBN: 9780521587211. DOI: `10.1017/CBO9781139174909.007` (cit. on p. 10).

[19] S. L. M. Heald and H. C. Nusbaum, Speech perception as an active cognitive process, *Frontiers in Systems Neuroscience*, vol. 8, 2014. DOI: `10.3389/fnsys.2014.00035` (cit. on p. 11).

[20] D. M. Jones, D. Alford, W. J. Macken, S. P. Banbury and S. Tremblay, Interference from degraded auditory stimuli: Linear effects of changing-state in the irrelevant sequence, *Journal of the Acoustical Society of America*, vol. 108, pp. 1082–1088, 3 2000. DOI: `10.1121/1.1288412` (cit. on pp. 12, 13).

[21] W. Ellermeier and J. Hellbrück, Is level irrelevant in ”irrelevant speech”? Effects of loudness, signal-to-noise ratio, and binaural unmasking. *Journal of Experimental Psychology: Human Perception and Performance*, vol. 24, pp. 1406–1414, 5 1998. DOI: `10.1037//0096-1523.24.5.1406` (cit. on pp. 13, 14).

[22] N. Venetjoki, A. Kaarlela-Tuomaala, E. Keskinen and V. Hongisto, The effect of speech and speech intelligibility on task performance, *Ergonomics*, vol. 49, pp. 1068–1091, 11 2006. DOI: `10.1080/00140130600679142` (cit. on pp. 13, 14).

[23] S. J. Schlittmeier, J. Hellbrück, R. Thaden and M. Vorländer, The impact of background speech varying in intelligibility: Effects on cognitive performance and perceived disturbance, *Ergonomics*, vol. 51, pp. 719–736, 5 2008. DOI: `10.1080/00140130701745925` (cit. on pp. 14, 15).

[24] V. Hongisto, A model predicting the effect of speech of varying intelligibility on work performance, *Indoor Air*, vol. 15, pp. 458–468, 2005. DOI: `10.1111/j.1600-0668.2005.00391.x` (cit. on p. 14).

[25] F. A. Muckler and S. A. Seven, Selecting Performance Measures: “Objective” versus “Subjective” Measurement, *Human Factors*, vol. 34, no. 4, pp. 441–455, 1992. DOI: `10.1177/001872089203400406` (cit. on pp. 14, 15).

[26] D. Kahneman, Attention and Effort. Prentice-Hall, 1973 (cit. on p. 15).

[27] M. K. Pichora-Fuller, S. E. Kramer, M. A. Eckert, B. Edwards, B. W. Y. Hornsby and L. E. H. et al., Hearing Impairment and Cognitive Energy: The framework for Understanding Effortful Listening (FUEL), *Ear and Hearing*, vol. 37, 5S–27S, 2016. DOI: `10.1097/AUD.0000000000000312` (cit. on p. 15).

[28] A. Kjellberg, Noise, in *Occupational Health Practice*, H. A. Waldron and C. Edling, Eds., 4th, Butterworth Heinenmann, 1997, pp. 241–256 (cit. on p. 15).

[29] J. Forssén, L. Müller, E. Hedlund and W. Kropp, Hybrid prediction tool implemented for acoustic design studies of open plan office spaces, *Applied Acoustics*, vol. 248, 2026. DOI: `10.1016/j.apacoust.2026.111284` (cit. on pp. 18, 20).

[30] H. S. Braren and J. Fels, “A High-Resolution Head-Related Transfer Function Data Set and 3D-Scan of KEMAR,” Institute, Chair for Hearing Technology and Acoustics, RWTH Aachen University, Tech. Rep., 2020. DOI: `10.18154/RWTH-2020-11307` (cit. on p. 21).

[31] P. Kabal, “TSP Speech Database,” Department of Electrical & Computer Engineering, McGill University, Tech. Rep. Version 2, 2018 (cit. on p. 21).

[32] IEEE Subcommittee on Subjective Measurements, IEEE Recommended Practice for Speech Quality Measurements, *IEEE Trans. Audio and Electroacoustics*, vol. AU-17, no. 3, pp. 225–246, 1969, (IEEE Standards Publication No. 297-1969). DOI: `10.1109/TAU.1969.1162058` (cit. on p. 22).

[33] K. Saberi and D. R. Perrott, Cognitive restoration of reversed speech, *Nature*, vol. 398, p. 760, 1999. DOI: `10.1038/19652` (cit. on p. 23).

[34] NASA Task Load Index (TLX): Paper-and-Pencil Version, NASA-Ames Research Center, Aerospace Human Factors Research Division, Moffett Field, CA, USA, 1986 (cit. on p. 26).