RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Doctorat de l'Université de Toulouse

délivré par l'ISAE-SUPAERO

Investigation et atténuation de l'effort d'écoute : approche électrophysiologique et comportementale.

Thèse présentée et soutenue, le 11 décembre 2025 par

Caroline HAMERY

École doctorale
SYSTEMES

Spécialité
Informatique

Unité de recherche
ISAE-ONERA DECISIO DECIsion, Supervision et Interaction pour l'Opération de systèmes complexes

Thèse dirigée par
Sébastien SCANNELLA

Composition du jury
M. Daniel PRESSNITZER, Président, CNRS-LSP
Mme Adriana ZEKVELD, Rapporteure, Amsterdam UMC
Mme Stefanie ENRIQUEZ-GEPPERT, Rapporteure, University of Groningen
M. Matthew WISNIEWSKI, Examinateur, Kansas State University
Mme Séverine SAMSON, Examinatrice, Université de Lille
M. Sébastien SCANNELLA, Directeur de thèse, ISAE-SUPAERO

Membres invités
M. Guillaume ANDÉOL, Institut de Recherche Biomédicale des Armées
M. Luc ARNAL, Institut Pasteur

Abstract

Oral communication is at the center of human interaction, and when the auditory scene becomes challenging, listening effort, defined as the "deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a task" (Pichora-Fuller et al., 2016), varies across individuals. In this project, we investigated listening in effortful auditory environments using both behavioral and electrophysiological approaches. Three experimental phases were conducted: (1) behavioral assessment of speech intelligibility and listening effort in native and non-native speech-in-noise and speech-in-speech conditions (N=51); (2) exploration of the relationship between executive functions and speech listening in challenging conditions, along with investigation of EEG alpha dynamics during effortful listening (N=30); and (3) cognitive training targeting inhibition to improve intelligibility and reduce listening effort, and evaluation of its effects on alpha dynamics (N=60). Behavioral results showed that using a native language enhances intelligibility and reduces listening effort, inhibitory control correlates with performance in the most adverse condition, and cognitive training improves speech perception while decreasing effort. EEG analyses confirmed the involvement of alpha oscillations with diverse neural generators during effortful listening. These findings emphasize the multidimensional nature of listening effort and its critical role in communication. Furthermore, they support the potential of cognitive interventions to mitigate listening challenges, with implications for clinical populations such as individuals with hidden hearing loss.

Acknowledgments

First, I would like to sincerely thank the members of the jury for accepting to be part of the defense of this thesis, and for their time and availability.

I am especially grateful to Adriana Zekveld and Stefanie Enriquez Geppert for reviewing the work presented in this manuscript.

My warmest thanks also go to Aurélie Bidet-Caulet and Anke Brock for their valuable advice and for kindly agreeing to be part of the thesis advisory committee.

C'est toujours une petite aventure, une thèse, et comme une aventure ne se vit jamais seule, je profite de l'écriture de ce manuscrit pour remercier les personnes qui ont vécu leurs propres aventures en parallèle à la mienne.

J'emporte avec moi la bienveillance, la gratitude et un petit pot rempli de chance, qui m'accompagneront tout au long de ma carrière et de ma vie. C'était une superbe aventure, il y en aura plein d'autres, et ça sera toujours grâce aux personnes qui les accompagnent.

Tout d'abord merci Seb, pour tout. Six années passent en un éclair. Alors merci pour les encouragements, la confiance et les blagues. Merci d'avoir fait de cette thèse une collaboration, merci de me montrer l'exemple et de me laisser tester ensuite. Merci de nous avoir fait sursauter en cognant la porte du bureau et ensuite d'avoir utilisé toute ton humanité pour comprendre et écouter.

Guillaume, merci pour toutes les portes que tu me présentes, elles sont toujours pleines de surprises. Merci de partager ta passion et ton énergie rebondissante avec ceux qui ont la chance de croiser ton chemin.

À toutes les personnes du labo, merci de faire de cet endroit un lieu qui vit et qui sonne comme une confortable maison. Merci pour votre présence chaleureuse, vos discussions scientifiques ou autres, et votre soutien dans les moments où je courrais partout et où on ne pouvait me parler que par créneau de 12 minutes. Merci Quentin pour les collab, la rigueur et les conseils. Merci Fred d'être venu chercher ton téléphone dans notre bureau, même s'il n'y était jamais. Seva, merci pour ton attention discrète, tes accès d'énergie et ton sourire en coin. Julien, merci pour les blagues et l'odeur d'agar-agar. Merci Laurence d'avoir passé ta petite tête tous les jours par la porte de notre bureau pour vérifier que tout le monde allait bien, même les plantes. Kalou, ta sincérité et tes conseils sont précieux, merci du temps que tu as pris pour m'aider et écouter. Inès ta joie et ta pétillance illuminent les lieux où tu passes.

Merci à toutes les personnes qui, en participant aux expériences de cette thèse, m'ont fourni du matériel précieux pour son écriture.

Aux amis de partout: Martine et Cedric, merci pour les fous rires sur les terrains de volley. Maxence, merci pour les dessins, ta douceur et ton amitié. Lucas merci d'avoir écouté et participé à des énumérations de bêtises. Aubin, merci pour ta bienveillance et les découvertes sous-marines glacées. L'équipe de roller derby, Rak, Wheely, Anna, Mozza et les autres, merci pour les bleus et l'adelphité. Caliani, bravo pour ta thèse, c'est une chance d'avoir nos routes qui se croisent régulièrement, en conférence ou autour d'un thé.

Claire, merci pour les ascenseurs émotionnels, les mots doux, ton regard bienveillant et ta présence rassurante.

Céline, je reste toujours pas trop loin pour te faire une coupe de cheveux rapide et t'attendre pendant des pauses pipis.

Émilie, merci d'adapter ton irrémédiable calme à mon dynamisme insensé. Merci pour l'amitié sincère.

Johanna et Jonas, avoir pu se connaître à travers différentes vies est unique, l'amitié reste, je vous garde tout le temps précieusement dans une petite poche dans mon cœur.

Merci à toutes les femmes inspirantes qui m'entourent de montrer tout ce qu'on peut faire, tout ce qu'on a fait et tout ce qu'on pourra faire.

Maëlis, Mathias, Marcel, avoir grandi à vos côtés pendant ces quatre dernières années a été un des plus beaux cadeaux de cette thèse, avec des modèles comme vous et une équipe aussi solide, on devient toujours plus fort.

Maë, avec toi la vie est plus folle, plus intense et beaucoup plus belle, ça colle bien à ton image. Partager un bureau autant de temps c'est connaître la signification de chaque soupir, respiration et même le rythme auquel tu grignotes des amandes va me manquer. Merci pour ton écoute, ton empathie, ta sincérité et surtout surtout surtout ton rire.

Mathou, personne n'a ta patience, merci d'être un conseiller, un psy, un ami, un joueur, un acrobate et de savoir discrètement retirer les poids quand tu en vois sur mes épaules. Avec toi, tout est moins dense. Bravo pour ta thèse, c'est une chance d'avoir vécu nos petites aventures doctorales en parallèle. Merci pour tous tes types de sourire, ils réchauffent toujours la pièce.

Marcel, il n'y a que toi pour créer des sentiments aussi incompatibles que l'admiration, la peur et l'affection en même temps. J'ai hâte de décoller dans un avion que tu piloteras, de stresser parce que je connais l'immensité de ta bêtise, et puis de me rassurer parce qu'on peut compter sur toi pour faire les choses comme il faut.

Anne-Sophie et Pauline, merci d'être entrées dans ma vie et d'avoir choisi d'y rester. Merci pour la sororité, d'être présentes, de comprendre qui je suis et de me donner accès à qui vous êtes.

Martin, merci pour la tendresse, les crampes aux joues et les petites boules à facettes que tu laisses dans chaque coin de ma tête et partout chez moi.

Merci à toute toute toute la famille pour l'amour force 1000.

À mes parents, merci de m'avoir offert deux frères et une famille solide et inspirante. Simon, Thomas, merci d'être là depuis toujours et pour toujours, vous êtes mes piliers.

Merci Papascalou, pour le soutien inconditionnel depuis toujours. Merci de t'être réveillé avant le soleil pendant 7 ans pour que je puisse m'échapper de l'autre côté du Rhin. Merci pour la transmission, la patience et la confiance.

Maman, merci d'être qui tu es, de n'avoir jamais douté de ce que je pouvais être et de m'avoir discrètement toujours donné les conseils les plus importants et les plus vrais. Merci pour l'inspiration, ta force, les discussions, les remises en question et d'avoir mis plein de petites étoiles dans mon cœur quand tu m'as fabriqué.

Merci aux absents, ceux qui sont passés et laissent de beaux souvenirs et ceux qui étaient là déjà avant nous et une fois disparus, restent en mémoire pour toujours.

Merci au magicien grâce à qui la vie est une histoire lumineuse et qui m'a toujours conseillé: “fonce, mais préserve-toi”.

Contents

Figures

Abbreviations

ANOVA	Analysis of Variance
ArsQ	Amsterdam Resting-State Questionnaire
CON	Cingulo-Opercular Network
CRM	Coordinate Response Measure
dB	Decibel
EEG	Electroencephalography
ELU	Ease of Language Understanding
EM	Energetic Masking
ERD	Event-Related Desynchronization
ERSP	Event-Related Spectral Perturbation
ERS	Event-Related Synchronization
ERP	Event-Related Potential
ESCU	Effort Scale Categorical Units
F0	Fundamental Frequency
fMRI	Functional Magnetic Resonance Imaging
fNIRS	Functional Near-Infrared Spectroscopy
FUEL	Framework for Understanding Effortful Listening
GFP	Global Field Potential
GEV	Global Explained Variance
HHL	Hidden Hearing Loss
IC	Independent Component
ICA	Independent Component Analysis
IFG	Inferior Frontal Gyrus
IM	Informational Masking
LE	Listening Effort
LEAP-Q	Language Experience and Proficiency Questionnaire
MD	Multiple Demand
MEG	Magnetoencephalography
MolE	Model of Listening Engagement
MS	Microstates
NASA-TLX	NASA Task Load Index
PET	Positron Emission Tomography
RAU	Rational Arcsin Unit
RSpan	Reading Span Task
RT	Reaction Time
SAS	Supervisory Attentional System
SI	Speech Intelligibility
SIN	Speech in Noise
SIS	Speech in Speech
SNR	Signal-to-Noise Ratio
SSQ	Spatial and Quality of Hearing Scale
STG	Superior Temporal Gyrus
TMR	Target-to-Masker Ratio
WM	Working Memory

I

Introduction

Oral communication is at the heart of human interaction, yet it can easily be affected by interferences that degrade its quality. Whether these disruptions come from the listener, the acoustic environment, or the speaker, they can create difficulties in understanding speech that vary from one individual to another. In the same situation, two people with similar hearing abilities may not need to invest the same amount of effort to understand a speaker. This specific effort, called listening effort, is the key concept of this PhD project. Although it is widely studied and described, listening effort remains difficult to measure, which makes its investigation challenging.

Hidden hearing loss, for example, affects people whose audiometric results appear normal but still experience auditory difficulties such as tinnitus, hyperacusis, or difficulty in understanding speech in noisy environments. Thus, listening effort varies between individuals and can sometimes become a burden, leading to social withdrawal. Therefore, studies aiming at defining, measuring and possibly mitigating this effort could have an impact on future clinical interventions.

Listening effort can be assessed in different ways, notably using self-reports or physiological measures. Using self-assessment methods, interpretations can vary from one individual to another, resulting in a subjective measure of listening effort, while physiological recordings may provide more objective data. Among the various physiological measurement tools, electroencephalography offers a non-invasive, high temporal resolution approach to measuring brain activity during listening. However, there is still no consensus on specific neural markers of listening effort, and each study contributes a new insight into this topic.

In this project, we chose to use electroencephalography to explore the neural correlates that may underlie listening in complex auditory environments. Different patterns of neural activity have been identified as potential indices of listening effort, such as alpha oscillations. The dynamics of these oscillations could be related to listening effort, and this project aimed to deepen our understanding of these mechanisms.

Therefore, we investigated listening effort from both behavioral and electrophysiological perspectives. First, an examination of the existing literature, exploring the auditory scene and how listening effort is defined (Chapter 1), is followed by an exploration of the neural pathways involved in listening (Chapter 2). Then, we explorer higher-level cognitive functions associated with speech understanding (Chapter 3), and finally we investigated cognitive training and its potential transfer effects to listening in complex auditory situations (Chapter 4).

The main objective of this project was to better understand different aspects of listening in effortful situations, particularly in relation to inhibition, a high-level cognitive function described as a key mechanism for the suppression of irrelevant or competing information. Inhibition may play an important role in complex auditory environments, especially when multiple speakers are present. We also investigated the possibility of mitigating listening effort with cognitive training, aiming to observe how such an intervention could affect challenging speech perception.

Based on these considerations, the following research questions and hypotheses were defined:

Questions and Hypotheses

Q1: Does the language used for a speech corpus influence listening in complex situations?

H1: The use of a native language would positively influence SI and LE, particularly in adverse situations.

Q2: Is there a relationship between listening in complex situations and inhibitory control?

H2: Executive functions performances (especially inhibition) would be correlated with speech intelligibility and with listening effort.

Q3: What are the neural correlates of listening in complex situations?

H3: Alpha oscillations dynamics would be impacted by the difficulty of the auditory scene.

Q4: Is it possible to improve listening in complex situations by training associated cognitive functions?

H4: A cognitive training of inhibition would improve SI and LE in multi-talker situations

Q5: What are the effects of cognitive training on the potential neural correlates of LE?

H5: Cognitive training would impact alpha dynamics differently depending on the trained cognitive process.

To address these hypotheses, three experimental phases were conducted:

Experimental phases

Language Behavioral measures of speech intelligibility and listening effort in native and non-native versions of the same corpus in speech-in-noise and speech-in-speech conditions. (N=51)

Executive Functions Investigation of the relationship between listening in effortful conditions and executive functioning. Exploration of alpha dynamics during effortful listening with an electrophysiological approach. (N=30)

Cognitive Training Implementation of cognitive training to mitigate listening effort in speech-in-speech conditions, using different training tasks. (N=60)

This thesis aims to investigate various aspects of listening in complex auditory environments, with a particular focus on its relation with inhibitory control, its mitigation, and the electrophysiological markers associated with these processes.

II

State of the Art

1

Complex Auditory Situations

Description of the Chapter

Auditory perception is a fundamental cognitive function that enables us to navigate and interact within complex acoustic environments. Listening in such environments can present challenges that depend on various factors related to the listener, the sound sources, and the environment itself. In this chapter, the auditory scene and the parameters contributing to its complexity are explored. Listening effort, which lies at the center of this project, is defined, and different approaches to its measurement are presented. Finally, the chapter addresses the importance of studying listening effort for real-life issues such as hidden hearing losses.

1 Complex auditory scene

1.1 The auditory scene

Like in a theater, a complex auditory situation can be imagined as a stage with different actors. This stage is the auditory scene. At the center of it is the main character: the listener, trying to gather and sort information from multiple sound sources (see Figure 1.1). To do so, the listener forms auditory objects by grouping and segregating sound streams. This process is highly complex, and although humans are remarkably good at extracting meaning from mixed signals entering their ears, the exact mechanisms behind this ability remain difficult to explain (Middlebrooks et al., 2017).

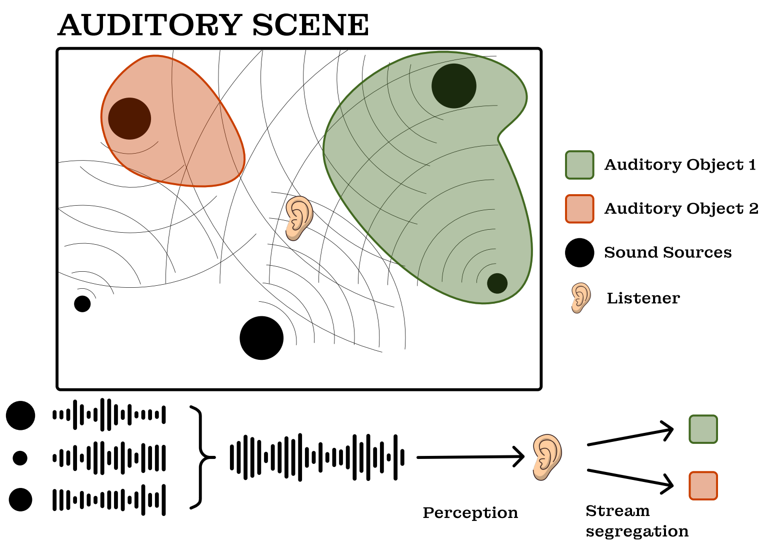


Figure 1.1: The auditory scene and auditory objects.

According to Middlebrooks et al. (2017), auditory objects refer to the mental representations of sound sources. The formation of an auditory object requires two key processes: grouping and segregating. Grouping brings together elements that likely originate from the same source, based on factors such as frequency, timing, harmonicity, or spatial location. Segregation, on the other hand, separates components from different sources. This can depend on differences in pitch, timbre, onset times, or spatial cues. Linguistic features, familiarity with voices, and even accents also influence how we segregate sounds (Johnsrude et al., 2013). Listeners analyze auditory objects based on knowledge and contextual information, engaging top-down cognitive processes (Davis and Johnsrude, 2007; see Box 1.5). According to Mattys et al. (2005), lexical and semantic cues derived from prior knowledge play a critical role in driving speech perception, showing the importance of top-down processes in interpreting complex auditory input. At any moment, the listener usually focuses its attention on a single auditory object, allowing selective perception within a complex auditory scene (Shinn-Cunningham et al., 2017).

Speech Chain

The speech chain, introduced by Denes and Pinson (1993), describes the flow of communication from the speaker's thoughts to sound production, transmission as sound waves, and perception by the listener. In the present manuscript, we briefly outline the physical properties of the sound wave as input to brain processes involved in speech perception (see Box 4. Speech chain impacts the acoustic challenge of a given auditory scene (Peelle, 2018).

1.2 Cocktail Party

In everyday life, we are frequently exposed to noisy and complex auditory environments, such as restaurants, supermarkets, train stations, or offices, where multiple sound sources compete for attention. The ability to separate auditory information is thus crucial and is widely studied. In 1953, Cherry described what has been called the "cocktail party effect" in a simple experiment with two simultaneous talkers. In multi-talker situations, this effect refers to the listener's ability to focus on the talker of interest to understand their message while ignoring others. This classic example of a complex auditory environment has since been widely studied in research on speech perception. The cocktail party problem occurs when competing sound sources are present in an auditory scene in which a listener tries to focus their attention on one stream of information.

Also, the cocktail party situation represents a dynamic environment with multiple overlapping conversations, often accompanied by non-speech noises such as music, construction sounds, or traffic. Despite the constant complexity of such auditory scenes, listeners continuously engage cognitive processes to maintain auditory focus, sometimes without conscious effort. However, when the complexity exceeds a certain level, comprehension difficulties arise. At some point, as auditory challenge increases, so do the engaged cognitive resources; this specific effort is called listening effort (LE) and is the topic of the next section.

2 Listening Effort

2.1 Description and definition

We tend to forget that our auditory system is constantly active, allowing us to hear and listen without conscious effort. However, the auditory scene can become more complex, making it difficult to understand speech. In such situations, the listener must adapt the exerted effort to the challenge. In addition, this LE has been shown to depend on several factors such as the environment (e.g. number of sound sources, background noise), the speech signal (e.g. sound degradation, unfamiliar talker, accent), and the listener's individual characteristics (e.g. auditory acuity, language proficiency) (Peelle, 2018). These factors make LE challenging to define, measure, and study. Although research on the topic is extensive and multiple authors suggest definitions according to their framework or gaze on the subject, there is currently no clear and consensual definition of it. Nonetheless, the set of existing definitions allows us to outline a conceptual framework.

Depending on the authors, LE may be defined from different perspectives. For instance, McGarrigle et al. (2014) described it as the ”mental exertion required to attend to, and understand an auditory message”. This definition, however, does not explicitly include concepts such as available cognitive resources, fatigue, or motivation. To address these limitations, Johnsrude and Rodd (2016) emphasize the interaction between the acoustic challenge and the cognitive resources available to the listener, highlighting that LE depends on multiple factors inherent to the listener, the speech specificity, and the environment. Considering that LE can be both exerted and perceived, Van Hedger and Johnsrude (2022) suggest a multidimensional view of LE as an ”interaction between the demands imposed by the listening situation and the unique constellation of cognitive abilities an individual listener brings to bear.” In their review, Francis and Love (2020) introduced an ”economic” view of LE, focusing on physiological markers, such as consumption of metabolites (e.g., glucose). They also distinguish between traditional active listening models and the FUEL model (Framework for Understanding Effortful Listening), which is one of the most widely cited and adopted definition of LE Pichora-Fuller et al. (2016). In this model, LE is described as a ”deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a task.”

Peelle (2018) pointed out the distinction between cognitive (or listening) demand, the challenges associated with the auditory scene and LE, and the resources actually used to meet the cognitive demand (see Figure 1.4). By stressing this difference, he underscored the influence of both external and internal factors on listening demand and suggested that it is, to some extent, a conscious choice made by the listener to exert LE. When the acoustic environment becomes more challenging, cognitive demand increases, leading to a greater LE that is modulated by the listener’s motivation.

B. Herrmann and Johnsrude (2020) suggested the use of the term ”listening engagement” that would include LE as a subjective experience. They argued that LE and engagement are two different processes.

Frameworks and definitions such as those of Peelle (2018), Pichora-Fuller et al. (2016), and Van Hedger and Johnsrude (2022) emphasize the multidimensionality of LE and the role of both internal and external factors. In each of these definitions, LE is viewed as a conscious allocation of resources by the listener.

Listening Effort

In the FUEL model, Pichora-Fuller et al. (2016) describe listening effort as *”deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a task.”*

2.2 Motivation, engagement, and effort

Listening effort is a multidimensional and dynamic concept influenced by both external and internal factors. Among internal factors, motivation is regularly highlighted in LE studies. Indeed, in complex auditory environments, a lack of motivation to attentively listen to a target talker can lead to disengagement and the absence of effective perception. Conversely, in situations with no auditory challenge, the perception and interpretation of speech occur automatically and effortlessly. According to the motivation intensity theory (Brehm and Self, 1989), the individual's capacity and the perceived success value determine the relationship between task demands and effort. When listeners perceive that the costs of listening outweigh the benefits, their motivation decreases, often resulting in reduced intelligibility (Matthen, 2016). This decrease in motivation can even lead to disengagement, where the listener ceases to exert effort to understand the speech signal.

To consider these factors, the FUEL (Pichora-Fuller et al., 2016) incorporates motivation within its model, applying motivation intensity theory to LE and highlighting motivation as a critical component.

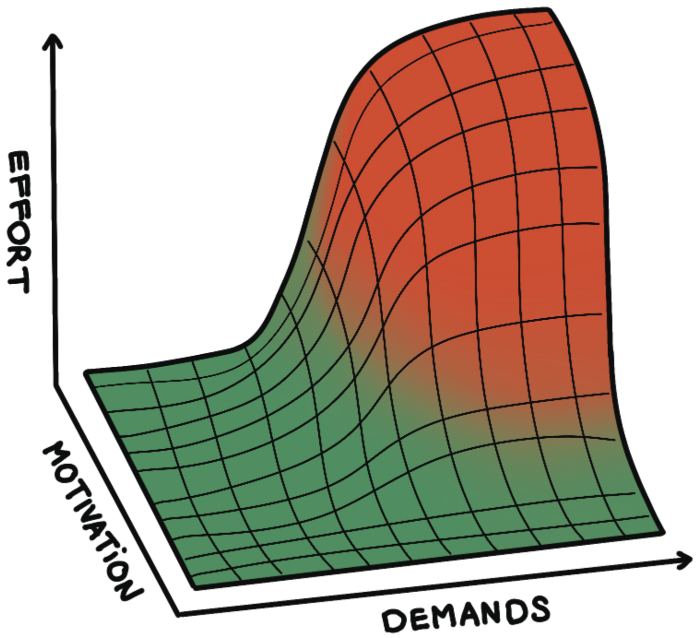


Figure 1.2: Framework for Understanding Effortful Listening Model (FUEL), illustrating effort variation as a function of the demands for capacity required for performance, and the motivation - adapted from Pichora-Fuller et al., 2016

B. Herrmann and Johnsrude (2020) further refine this understanding through their model of listening engagement (MolE, see Figure 1.3), which conceptualizes listening engagement as the recruitment of executive and cognitive resources, whether automatic or deliberate, to achieve a valued understanding goal. This model suggests that motivation, engagement, and listening experiences are deeply interrelated, and that research should target these interactions to better understand LE. Moreover, it assumes that listeners possess a variable (flexible) pool of cognitive resources that can be adapted to meet the demands of their listening environment, underscoring the active and conscious role of listeners in managing effort.

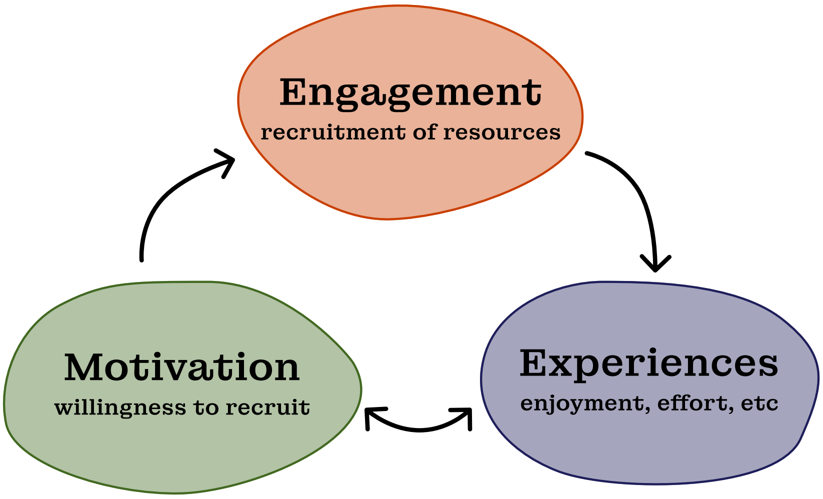


Figure 1.3: Model of listening engagement (MolE), illustrating the relation between engagement, motivation, and experiences - adapted from B. Herrmann and Johnsrude, 2020

Listening effort is often described as the amount of engagement a listener applies to the task; when task demands increase, greater effort is required, leading to increased effort investment (Richter, 2016). This distinction between listening demand (task difficulty) and LE (amount of mobilized cognitive resources) underscores that motivation plays a central role. The very use of terms such as ”investment”, ”cost”, or ”benefits” reflects the idea that motivation is an important underlying factor that affects both LE and speech intelligibility.

By distinguishing between the subjective experience of listening difficulty and the cognitive processes involved in effortful listening, the MoLE helps clarify some confusion regarding the various definitions of LE in the literature. Although some authors define LE as the subjective perception of difficulty (Johnsrude and Rodd, 2016; Krueger, Schulte, Zokoll, et al., 2017), others view it as the mental act of investing effort (McGarrigle et al., 2014; Peelle, 2018; Pichora-Fuller et al., 2016). The MoLE model shows how both conscious and automatic aspects of engagement help achieve intelligibility goals.

Furthermore, disengagement in speech perception remains understudied (B. Herrmann and Johnsrude, 2020); nevertheless, it is common to observe patterns in experimental data that suggest that participants disengaged when listening conditions get too challenging.

Therefore, it is important to consider the motivational context when designing and interpreting LE studies (Carolan et al., 2022).

For more details on the relationship between LE and motivation, a very detailed review by Carolan et al. (2022) outlines five categories of motivation in experimental LE research: financial reward, evaluative threat, perceived competence, feedback, and individual traits.

2.3 Speech Intelligibility

Conceptually intertwined with LE, speech intelligibility (SI) refers to the objective measure of how accurately a listener can understand a target speech. It is widely used to evaluate speech perception, particularly in complex auditory environments. Unlike LE, which can be self-reported, SI is typically assessed through performance-based tasks, often quantified as the percentage of correct responses.

It has been shown that SI performances and LE are not always related and can be independent (Strauss and Francis, 2017; Winn and Teece, 2021). Thus, the amount of cognitive effort a listener allocates does not necessarily predict their intelligibility. In some conditions, speech may be highly intelligible with minimal effort; in others, even intense effort may not yield an accurate understanding of the speech. This exposes the importance and challenge of studying both constructs in parallel.

According to Winn and Teece (2021), using intelligibility scores as a proxy for LE is "tempting and intuitive" but oversimplifies the complexity of cognitive load in speech perception. Individuals with poorer intelligibility scores are not necessarily experiencing greater LE and laboratory material and methods may not reflect the full demands of natural language processing. Strauss and Francis (2017) emphasize that LE should be investigated as a distinct construct from SI.

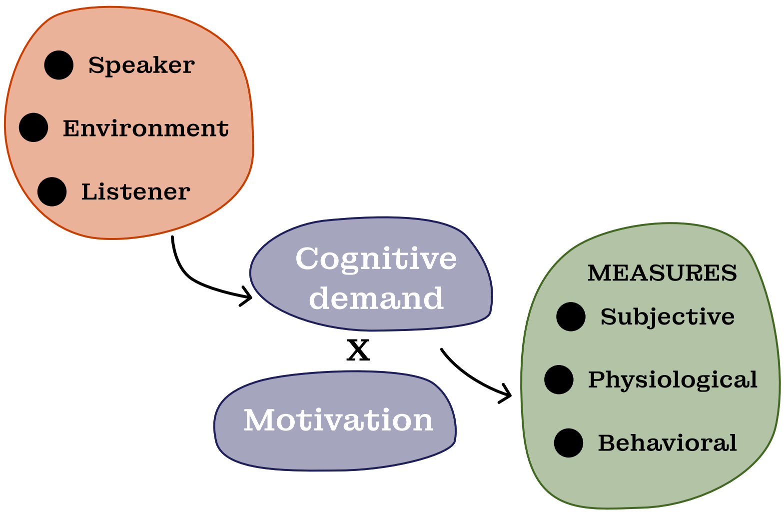


Figure 1.4: Actors influencing cognitive demand in the auditory scene - adapted from Peelle, 2018

2.4 Individual differences

Considering the aforementioned factors, individual differences may also result in different LE. For instance, listeners with similar hearing thresholds may exert different levels of LE to achieve comparable intelligibility. This inter-individual variability lies at the core of this thesis project. As a consequence, the main question to address this issue could be stated as follows: why do some individuals cope better than others in the same auditory conditions?

Hearing impairments are known to increase LE and reduce SI (Peelle, 2018; Strand et al., 2018), even in auditory scenes comparable to those experienced by individuals with normal hearing (Bess and Hornsby, 2014). Hearing devices such as hearing aids can help reduce this effort in certain conditions, depending on the type and severity of the impairment (Holman et al., 2021). However, LE and SI still vary significantly among listeners in adverse conditions, regardless of hearing status.

Therefore, pure tone audiometry is not sufficient to represent or describe the human hearing abilities on its own, and some authors Akeroyd (2008) and Peelle (2018) suggest that cognitive abilities have a role in individual differences of SI and LE in complex auditory situations. In other words, the cognitive demand required in a given auditory situation depends on a range of internal and external factors that are not fully captured by pure tone audiometry.

One question arises from these observations: What are the reasons for these differences? In a review of 20 studies, Akeroyd (2008) showed that the impact of hearing loss on SI was more important than the cognitive abilities. However, this does not announce that the link is the same between LE and cognitive abilities.

Inter-individual differences in LE may also be explained by brain mechanism processes. Francis et al. (2021) proposed that physiological measures may reflect how effectively listeners engage cognitive systems, and that individual differences in these capacities could help explain the variability in LE.

Studies often emphasize the need for objective and reliable markers of LE. Such indices would enable more personalized diagnostics, improved care, and better clinical support for individuals struggling in complex auditory environments. An increasingly challenging condition may lead to more cognitive load and, with it, the increase of inter-individual differences. Understanding the underlying cognitive and neural factors driving these differences is therefore useful.

Top-down and Bottom-up Processes

Perception can be influenced by two distinct processing approaches: bottom-up and top-down. The bottom-up approach is data-driven, meaning perception directs cognition. It relies on raw sensory input without the influence of prior knowledge or expectations, guiding decision-making and behavior based solely on external stimuli. In contrast, the top-down approach involves conceptually-driven processing, where conditions shape perception. Here, prior knowledge, experience, and expectations influence how sensory information is interpreted, meaning that perception and behavior are largely determined by cognitive frameworks and conceptual understanding.

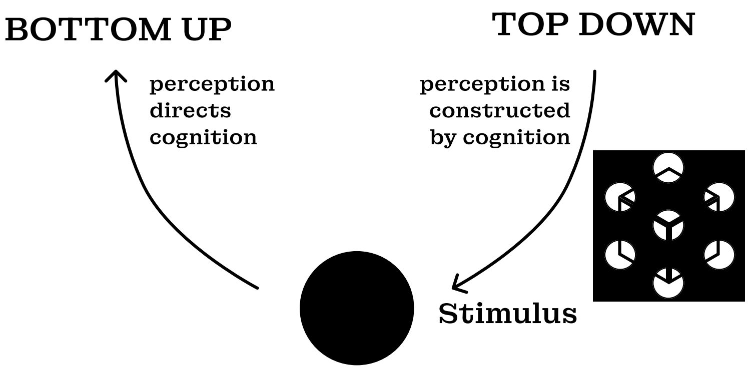


Figure 1.5: Top-down and Bottom-up processes

3 Measures of Listening Effort

As mentioned earlier, LE is a multidimensional concept. The complexity of its definition and the variety of perspectives in the literature highlight the need to align the measurement approach with the chosen definition. Of course, the measure of LE depends on the definition we decide to keep. When studying such cognitive processes, providing a clear definition of the concept is a step that will strongly impact the discussions around the results.

Furthermore, different types of measures do not capture LE in the same manner, and studies have shown that they rarely correlate (Alhanbali et al., 2019; B. Herrmann and Johnsrude, 2020; Miles et al., 2017; Pichora-Fuller et al., 2016; Strand et al., 2018). This problem can be explained by the multidimensional nature of LE and the variety of cognitive processes it encompasses (Pichora-Fuller et al., 2016); each type of measure is influenced differently by different aspects of LE.

Overall, it has been shown that LE can be measured through different methods such as self-reports, behavioral, and physiological measures (see Table 1.1 for a summary). However, there is currently no consensus on the right way to measure LE. In this section, we will describe the most commonly used measuring methods in the literature.

As Peelle, 2018 mentioned, LE is, in certain aspects, such as the related cognition, measurable. This measure depends on the acoustic challenge that occurs in the auditory scene, such as those discussed in the section 4.

3.1 Behavioral measures

Measures of SI in complex auditory situations are often considered alongside LE assessments. However, it is important to clarify that behavioral tasks report objective performance in SI and do not directly assess LE. These tasks can be investigated in parallel with self-reported LE ratings and physiological measures, but do not provide direct insight into the cognitive resources allocated during listening. In the context of LE research, various approaches exist to measure SI.

Generally, SI is assessed in controlled environments simulating noisy or multi-talker situations. These experimental conditions can be manipulated using a range of acoustic parameters, including noise type, source degradation, source localization, level differences between target and masker or number and type of talkers.

Some studies use a dual-task paradigm to increase cognitive load and related LE. However, based on the definition of LE adopted in the present project, such methods are not directly aligned with our objectives. For a detailed review on dual-task paradigm for the study of LE, please refer to Gagné et al. (2017).

SI can be assessed through several metrics, such as response time, percentage of correct responses, or speech reception thresholds. In natural language and everyday conversation, intelligibility is difficult to measure due to the variability of listener interpretations, among other factors. To address this matter, standardized speech corpora are often employed in experimental settings. They help minimize semantic ambiguity, focus solely on acoustic perception, and improve reproducibility across experimental protocols.

Among the very large pool of tests used in speech research, a few commonly used corpora in SI and LE studies are described below. To reduce semantic predictability and memorization effects, many studies use structured corpora such as matrix sentence tests with fixed syntactic patterns and semantically unpredictable context (see

Kollmeier et al., 2015 for a review). Examples include the Oldenburg Sentence Test (OLSA) in German (Wagener et al., 1999; e.g., used by Hall et al., 2019), the Russian Matrix corpus (Warzybok, Zokoll, et al., 2015) or the Danish Dantale II data base (Wagener et al., 2003; e.g., used by Mohammadi et al., 2023). More natural sentence corpora, such as the Hearing in Noise Test (HINT; Nilsson et al., 1994), TIMIT (Zue et al., 1990; e.g., used by Horton et al., 2013) or the Bamford-Kowald-Brench (BKB; e.g., McMahon et al., 2016; Miles et al., 2017), are also commonly employed. Additionally, some corpora are based on principles of military communication, such as the Modified Rhyme Test(MRT; House et al., 1963) or the Coordinate Response Measure (CRM; Bolia et al., 2000, adapted from T. J. Moore, 1981), which has been used by Lanzilotti et al. (2022), Brungart (2001b),Brungart (2001b), Wisniewski et al. (2021), Hamery et al. (2023)or Hambrook and Tata (2019) among others.

SI tasks are widely used to assess hearing disorders and, more generally, auditory abilities of participants in challenging listening situations (DiNino et al., 2022). Comparative analyses across experimental and application contexts rely on tasks that are configurable, particularly with respect to participant characteristics and task parameters (e.g., language, spatial distribution of attention), as well as the type of masking (e.g., energetic or informational). Among the tools used to study speech intelligibility, the Coordinate Response Measure (CRM) corpus (Bolia et al., 2000), adapted from T. J. Moore (1981), is widely used, due to its simplicity and adaptability (Brungart, 2001b; Mesgarani and Chang, 2012; Wisniewski et al., 2021). Like other intelligibility tasks, it enables researchers to assess speech comprehension in different complex auditory scenarios.

Among existing tools, an advantage of using a context-free corpus such as the CRM corpus lies in the possibility of isolating acoustic and perceptual processing by removing semantic cues. This helps ensure that the observed changes in performance are due to experimental conditions rather than top-down linguistic predictions (Bolia et al., 2000). Indeed, when all sentences follow the same pattern and structure, this facilitates the creation of controlled energetic and informational masking conditions.

Many of these corpora have been adapted in various languages; for example, the HINT has been translated into Cantonese by Wong and Soli (2005) and Swedish by Hällgren et al. (2006), the MRT in French (Zimpfer et al. (2020)) or the CRM in French by Isnard et al. (2024).

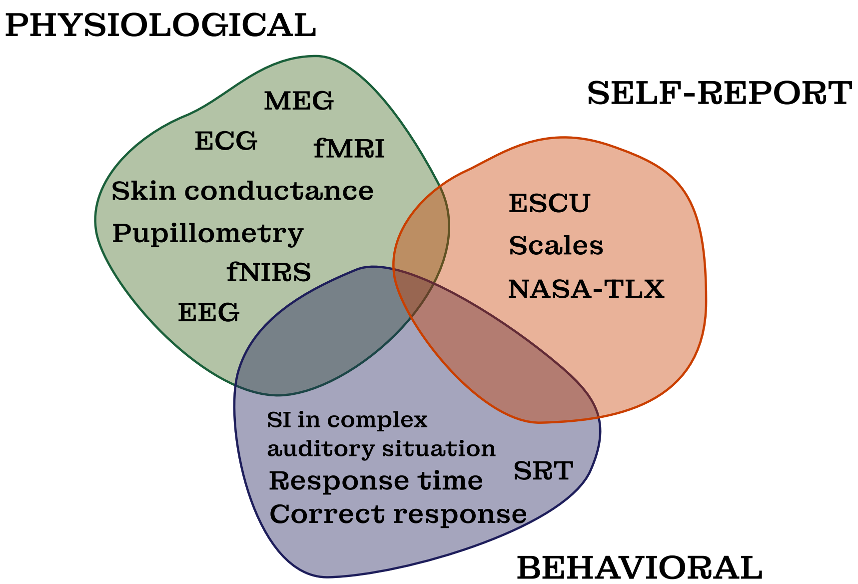


Figure 1.6: Listening effort measures representing its multidimensionality - adapted from Shields et al., 2022. SRT: speech reception threshold. SI: speech intelligibility

3.2 Subjective measures of LE

No standardized objective behavioral task for measuring LE currently exists. Consequently, behavioral assessments often rely on subjective scales or questionnaires. However, studies tend to show that LE can be objectively assessed with physiological measures (see in Section 3.3).

Based on a subjective scale used by Luts et al. (2010), Rennies et al. (2014) introduced the Effort Scaling Categorical Units (ESCU) for the self-report of LE. This scale is composed of 13 levels, including 7 categories labeled from ”no effort” to ”extreme effort” and 6 non-labeled intermediate categories. A fourteenth additional category, ”only noise” (or ”I cannot understand the target talker at all” in Krueger, Schulte, Brand, and Holube, 2017; Rennies et al., 2019), is often provided for the case in which listeners cannot understand anything, helping to adapt the parameter range.

An adapted version of this method, the Adaptive Categorical Listening Effort Scaling (ACALES; Krueger, Schulte, Brand, and Holube, 2017) was developed to reduce bias caused by random stimulus presentation and individual differences. ACALES controls each response based on previous ratings; depending on the listener’s personal experiences and expectations, individual differences could affect the results (McGarrigle et al., 2014).

Other questionnaires, such as the Speech, Spatial and Qualities of hearing scale (SSQ; Gatehouse and Noble, 2004) are used to assess LE in complex auditory situations. Originally designed for the measure of hearing impairments, the SSQ contains questions to evaluate difficulties in hearing and listening across various realistic acoustic environments. Besides, some studies use the NASA Task Load Index (NASA-TLX;Hart and Staveland, 1988), which was developed to measure task load and is used with slight modifications for LE assessment.

As noted by T. M. Moore and Picou (2018), subjective ratings of LE are less precise than physiological or behavioral measures. Also, directing listeners' attention to their own internal states may modulate their representation of the task, potentially reflecting how effortful they believe the task should be rather than how effortful it actually was (Francis and Love, 2020). This makes it complex to distinguish whether the listeners are reporting exerted or expected LE. Francis et al. (2021) further support this idea, suggesting that when listeners are asked to evaluate their exerted effort, they may actually be reporting a different aspect of their LE experience. Nevertheless, subjective ratings provide valuable insights into the listener's personal experience of effort during the task.

3.3 Physiological measures of LE

Objective assessment of LE often relies on physiological measures, which aim to identify neural and autonomic markers associated with effortful listening. These measures can reflect activity from both the central and peripheral nervous systems.

Neuroimaging techniques such as functional Near-Infrared Spectroscopy (fNIRS), Positron Emission Tomography scan (PET), functional Magnetic Resonance Imaging (fMRI), magnetoencephalography (MEG) and electroencephalography (EEG) allow observation of brain activity during listening tasks in complex auditory environments. Analyses of these signals provide insights into the neural mechanisms underlying auditory processing under challenging conditions.

In addition to these central measures, peripheral physiological responses, such as pupil dilation, skin conductance, and cardiac activity, also offer indirect insight into the cognitive demand of listening. These measures are believed to reflect modulation by central nervous system activity and are frequently used in combination with neuroimaging of behavioral data to provide a more comprehensive assessment of LE.

fMRI

fMRI studies have revealed key neural networks activated during LE situations. For instance, Eckert et al., 2016 conducted a meta-analysis showing consistent involvement of the cingulo-opercular network, including bilateral dorsal cingulates, inferior frontal and anterior insular cortices, during challenging but intelligible conditions. This network appears to play a critical role in top-down control and maintaining task engagement under degraded listening scenarios (Francis and Love, 2020).

Electrophysiology

Speech perception in adverse conditions is widely studied in EEG. However, the associated LE is not always explicitly examined. Given the conceptual distinction between SI and LE, their respective neural correlates are likely to differ. Nevertheless, it is sometimes unclear whether authors are referring to SI, LE or both.

EEG provides a flexible and sensitive tool for investigating the brain's response to complex auditory scenes and appears well-suited to assess the neural correlates of LE. Commonly used analytical methods include frequency, time, and time-frequency domains. Other approaches include microstate analysis, component analysis, and more advanced methods such as source localization, neural or cortical tracking, clustering, and deep learning techniques.

There are multiple ways to use EEG signals for analysis. Regarding speech perception, these methods can be applied in the time domain with ERP (Bertoli and Bodmer, 2014; Billings et al., 2009; Obleser and Kotz, 2011; O'Sullivan et al., 2015; Power et al., 2012; Wisniewski, 2017) or more precisely auditory ERP (Billings et al., 2009; Choi et al., 2013; Lewald and Getzmann, 2015; Papesh et al., 2017; Power et al., 2012; Wisniewski et al., 2023). Also, analysis can be conducted in the frequency domain (Giraud and Poeppel, 2012; Hall et al., 2019; Hambrook and Tata, 2019; Hunter, 2020; Obleser et al., 2012; Paul et al., 2021, or in the time-frequency domain with ERPS (Alhanbali et al., 2019; Nourski et al., 2009; Wisniewski et al., 2015, 2017). And in addition, microstates (Eqlimi et al., 2023; Roushan et al., 2023), components analysis (Jenson, Bowers, et al., 2014; Wisniewski et al., 2015, 2017, 2021, 2024, source localization (Wisniewski et al., 2024) or more complex concepts such as neural or cortical tracking (Ershaid et al., 2024; Hambrook and Tata, 2019; Horton et al., 2014; Hunter, 2020; O'Sullivan et al., 2015) or other classifier and deep learning procedures (Ding and Simon, 2012, 2014; Mesgarani and Chang, 2012; Puffay et al., 2024; Tian et al., 2018) can be used to analyze EEG data.

Concerning spectral activity, LE has been associated with modulations in alpha (8-12 Hz) and theta (4-8 Hz) power. Alpha rhythms, oscillating between 8 and 12 Hz, have received significant attention due to their involvement in various cognitive processes. Alpha synchronization has been linked to the suppression of irrelevant noise or sound, while desynchronization may reflect increased attentional focus on speech signals (Dimitrijevic et al., 2017; Obleser and Kotz, 2011; Strauß et al., 2014; Wilsch et al., 2015; more details in the next Chapter 2). Both increases and decreases in alpha power can co-occur in separate brain areas, potentially reflecting different aspects of LE (Paul et al., 2021).

Some evidence for alpha as a neural marker for LE shows that its power increases with greater acoustic degradation during speech listening tasks (Obleser et al., 2012; Wilsch et al., 2015). These enhancements are often stronger over temporal and occipital regions and may reflect the suppression of irrelevant or distracting information. Alpha power increases have also been noted in the primary auditory cortex, suggesting a role in inhibiting the formation of competing auditory objects (Leske et al., 2014; Weisz and Obleser, 2014). Some of these findings have been observed in non-speech tasks, suggesting that they may not be directly related to speech-specific processes (Wisniewski, 2017). While these studies suggest an increase in alpha power in relation to listening in complex environments, the role of alpha waves in speech intelligibility remains unclear, with various authors offering different interpretations.

Alpha waves have been extensively explored in the context of speech intelligibility and listening effort. However, there is still no clear consensus on the exact relationship between alpha activity and these factors. Alpha power desynchronization, meaning reduction of alpha amplitude, has been associated with increasing listening effort in complex auditory situations (Ala et al., 2023; Dimitrijevic et al., 2017; Hall et al., 2019. This desynchronization may reflect greater neural engagement and attentional resource allocation.

Wisniewski and Zakrzewski, 2023 investigated how alpha power varies during speech perception in a complex auditory environment, specifically in a multi-talker situation. They found that increased LE was associated with both alpha enhancement and suppression, depending on the subsequent alpha generator involved.

Specifically, left somatomotor mu rhythms showed enhancements, while left temporal tau rhythms exhibited suppressions, suggesting that different cortical sources contribute differently to listening effort. This dual modulation could explain inconsistencies in earlier studies that reported either increased or decreased alpha power during effortful listening.

These findings underscore the complexity of neural mechanisms underlying LE and highlight the importance of considering multiple brain regions and their interactions when interpreting EEG data in auditory tasks. In addition, Wisniewski and Zakrzewski, 2023 EEG analysis methods use more complex EEG activation patterns than simple spectral or time analyses. The component analysis procedure will be introduced in a later section and subsequently applied to our data.

In parallel to alpha waves, theta oscillations, especially the frontal midline theta, observed in the frontal regions, are often associated with cognitive effort (Francis and Love, 2020). Frontal midline theta power increases have been interpreted as markers of cognitive control and working memory engagement during effortful listening (2015, 2017). Wisniewski et al. (2017) demonstrated increases in both alpha and theta power bands relative to increased difficulty and decreases when the task becomes too difficult.

Despite these advances, the interpretation of electrophysiological markers remains complex. Some indicators (e.g., alpha power, tau components) may reflect different or even opposite mechanisms depending on the task or cognitive demand. Further research, with source localization, for example, is needed to disentangle these processes and define reliable neural indices of LE (Francis and Love, 2020).

Pupillometry

Among peripheral measures of LE, pupillometry is one of the most widely used methods (Zekveld et al., 2018). It involves measuring changes in pupil diameter in response to task-related events, such as auditory stimuli.

Indeed, as shown in several studies, pupil dilation is considered a reliable index of LE (McGarrigle et al., 2014; Zekveld et al., 2018). Most of them have shown that pupil size increases with higher LE or under conditions of greater sound degradation Koelewijn et al., 2012; Kramer et al., 2016; Miles et al., 2017; Zekveld et al., 2010, 2018.

In a review on pupil dilation response to auditory stimuli, Zekveld et al. (2011) noted that, in accordance with the FUEL model (Pichora-Fuller et al., 2016), pupil dilation is influenced by several factors, including the characteristics of the sound, the type of degradation, and individual listener differences. They concluded that pupillometry provides a robust and replicable measure of LE. However, because LE is a multidimensional construct, pupil dilation may reflect only certain of its underlying components (Strand et al., 2018). As such, it is often recommended to combine pupillometry with performance-based and/or other physiological measures for a more comprehensive assessment of LE.

Other physiological measures

In addition to neuroimaging and pupillometry, other peripheral physiological measures have been explored to assess LE, although findings across studies remain mixed. Among these peripheral indicators, skin conductance, or electrodermal activity, reflects autonomic nervous system arousal, specifically sympathetic activation. It tends to increase under cognitively or emotionally demanding conditions, making it a candidate marker for LE during challenging listening tasks (Alhanbali et al., 2019; Mackersie and Cones, 2011). However, it is also highly sensitive to emotional states, which can introduce variability and reduce its specificity for LE (Alhanbali et al., 2019; Hogervorst et al., 2014). In addition to skin conductance, cardiac measures have also been investigated. Some studies suggest they may reflect cognitive effort during listening, but results remain inconsistent (Francis and Love, 2020; Mackersie and Cones, 2011; Mackersie et al., 2015; Winn et al., 2018).

Finally, neurochemical measures, such as changes in cortisol, aldosterone, epinephrine, and norepinephrine levels, have been proposed as potential indicators of LE via activation of the hypothalamic-pituitary-adrenal and sympathetic-adrenal medullary axes (Francis and Love, 2020). However, these approaches are more invasive and fall outside the scope of the present project.

Measure	**Advantage**	**Limitations**
Self-assessment	Quick and easy to administer captures listeners' perception of LE; can be tailored for momentary (e.g., trial-by-trial) or global ratings	Subjective, influenced by individual interpretations, expectations, personality wide inter-listener variability
Behavioral: Single task	Direct measure of task performance. Easy to implement	Not a LE effort measure, is useful in parallel with other measures
Behavioral: Dual-task paradigms	Sensitive to extra processing load can reveal hidden costs not seen in primary task performance	May not reflect natural listening
Pupillometry	Good temporal resolution. Reproducible effects	Sensitive to individual's state, large inter-individual variability
EEG	Millisecond temporal resolution. Sensitive to neural dynamics (e.g., alpha/theta modulations) linked to LE. Relatively inexpensive and portable	Poor spatial resolution. Multiple analysis approaches. Susceptible to muscle and eye artifacts
fMRI	High spatial resolution. Maps cortical/subcortical networks involved in LE	Noisy, expensive low temporal resolution. All implicated networks have multiple functions
fNIRS	Portable Easy to install and handle. Good spatial resolution	Lower spatial resolution than fMRI Signals sensitive to scalp blood flow changes

Table 1.1: Summary of listening effort measures with advantages and limitations

3.4 Inconsistency across measures and Reliability

The multidimensional nature of LE makes it both accessible and challenging to assess. It is accessible because various aspects of LE can be measured using a wide range of tools and methodologies. However, it is also complex due to the diversity of these aspects and the many ways results can be interpreted.

Inconsistencies and lack of reproducibility are frequently reported in studies investigating listening in complex environments (Alhanbali et al., 2019; Keur-Huizinga et al., 2024; Lau et al., 2019; Strand et al., 2018). These inconsistencies may arise from individual differences in physiological reactivity or from methodological variations across studies. Research protocols may differ in terms of task type (e.g., sentence matrices, word corpora, digits in noise, audio books), listening conditions (e.g., speech-in-speech, speech-in-noise, non-speech stimuli, babble, noise type, spatial configuration, or source degradation), population (e.g., young healthy-hearing, hearing impaired, older adults, or non-native listeners), or even in the signal processing techniques used.

Furthermore, different LE measures do not always produce consistent results. For example, physiological indicators of LE often show weak or no significant correlations with each other or with subjective reports (Lau et al., 2019; Strand et al., 2018). As highlighted by Alhanbali et al. (2019), inconsistencies occur not only between measures but also between participants, groups, and studies using the same measures but different stimuli.

There are several other explanations for this variability. Some measures may lack sufficient reliability, making it unlikely for them to correlate with one another. In other cases, correlations might only appear when studies involve the same listening tasks, suggesting that methodological consistency plays a crucial role (Alhanbali et al., 2019). Additionally, subjective, behavioral, and physiological measures may each reflect distinct underlying constructs related to LE (Pichora-Fuller et al., 2016). This highlights the multidimensional nature of LE and suggests that no single method can capture all aspects of it.

In conclusion, the complexity of LE requires the use of multiple complementary assessment methods to account for its different aspects and underlying processes.

4 Adverse conditions

In real-world settings, the auditory scene often gets challenging for listeners, and conditions for understanding speech are far from optimal. Factors that decrease SI are referred to as adverse conditions (Andéol et al., 2017; Lanzilotti et al., 2022; Mattys et al., 2012). In addition, these conditions generally do not arise from a single cause. Instead, they often originate from different causes related to the three actors of the auditory scene: the speaker, the environment, and the listener.

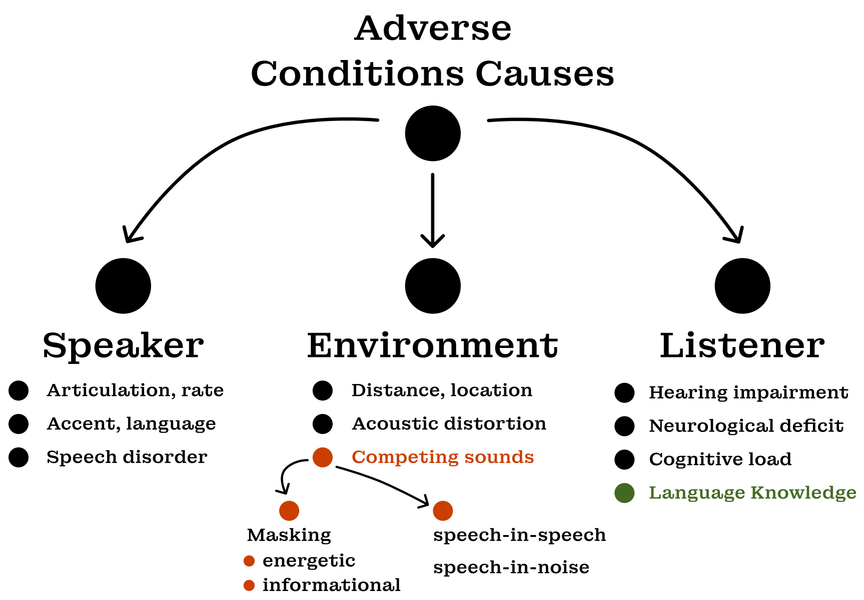


Figure 1.7: Adverse conditions causes.

These various sources of degradation can be categorized based on their origin (see Figure 1.7). One important type of degradation involves speaker-related adverse conditions (source degradation), which refer to characteristics of the talker that reduce intelligibility. This includes unclear articulation, speaking rate, accent, speech disorder or simply unfamiliarity (Johnsrude et al., 2013; Mattys et al., 2012; Van Hedger and Johnsrude, 2022).

Environmental conditions, causing transmission degradation, involve external factors that interfere with the speech signal. The transmission can be affected by factors such as the distance and orientation between the speaker and listener, acoustic distortions, reverberations, or competing sounds in the environment (Mattys et al., 2012). Competing signals result in energetic masking, informational masking, or both (see Section 4.1 for details).

Finally, listener-related factors, due to the receiver's limitations, can also impact speech perception. For instance, hearing impairments, neurological disorders, cognitive load, or incomplete language knowledge will affect intelligibility.

Also, these different types of adverse conditions can combine. In most real-world situations, listeners do not face just one challenge but rather a complex mixture of several. For instance, a person might try to understand speech in a second language in a noisy environment while also being distracted or tired. As explained by Van Hedger and Johnsrude (2022), different types of adverse conditions place different demands on cognitive resources, and their interaction can significantly increase the LE required.

Sound physics

Sound is a mechanical wave caused by the movement of air molecules, generated by the vibration of an object. The resulting pressure variations cause the molecules to move back and forth from their resting position. A sound wave is characterized by two key features: frequency and amplitude. The frequency, defined as the number of cycles occurring per second, is measured in hertz (Hz) and determines the pitch of the sound. The human ear can perceive frequencies ranging from 20 to 20 000 Hz. The amplitude refers to the intensity or loudness of the sound; it corresponds to the maximum movement of the molecules. When illustrated as in Figure 1.8, the amplitude of a sound corresponds to its pressure, expressed in Newtons per square meter (N/m^2). The decibel (dB) scale, used to describe sound intensity, is logarithmic, meaning that for every increase of 10 dB, the sound intensity becomes ten times greater.

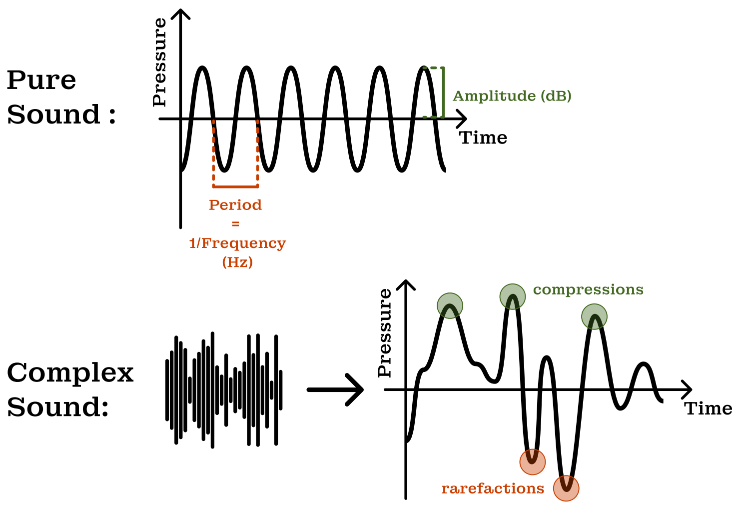


Figure 1.8: Pure and complex sounds physics.

Pure sounds, such as illustrated in Figure 1.8, rarely occur in natural conditions. Most of the sounds we hear in everyday life are complex, consisting of multiple frequencies and composed of alternating compression and rarefaction phases, corresponding to changes in air pressure caused by the air molecules' movements. The sound propagates in three dimensions, meaning it travels in all directions from its source.

4.1 Energetic and Informational Masking

Masking occurs when the perception of one sound is impacted by the presence of another. The masking effect can operate at different levels. First, energetic masking (EM) occurs when a competing sound interferes with the target one in time and frequency (Bronkhorst, 2015; Kidd and Colburn, 2017). The masking of the stream of interest is due to the mix of energy of the two streams at the periphery of the auditory system (Durlach et al., 2003). Second, informational masking (IM) is considered the result of the masking signal once the effects of energetic masking have been compensated for (Cooke et al., 2008). IM engages higher-level cognitive resources and occurs at the central level of auditory processing, where competing speech-like signals overlap with the target in the speech spectrum and semantically interfere with its interpretation (Bregman, 1990; Brungart, 2001a; Culling and Stone, 2017).

Masking factors not related to EM are considered IM, for example, when other speech signals interfere with the target speech (Kidd and Colburn, 2017). In a cocktail party, masking makes it harder for the listener to extract auditory objects from the auditory scene. To overcome this masking, the listener employs release strategies; the process by which the target signal (in our case, speech) is separated from the competing sounds, thereby improving SI.

Research studies often manipulate masking conditions to investigate LE and SI. However, in speech-in-speech (SIS) situations, it is difficult to completely separate EM and IM since both the masker and the target are speech signals. In contrast, speech-in-noise (SIN) scenarios are often used to isolate EM, as the noise maskers do not contain semantic information. Determining the proportion of EM and IM in SIS is challenging. An experimental approach to studying IM while minimizing this bias is to keep EM constant (Kidd and Colburn, 2017).

The difference in production between EM and IM is also represented at different processing levels in the brain. EM is mainly handled by a lower, peripheral level of the auditory system, where the target signal is physically blocked by a masker that contains no semantic content (Bronkhorst, 2015). The auditory system, however, is capable of using various strategies to overcome such masking and extract relevant information (Culling and Stone, 2017).

4.2 Masking release strategies

Masking, whether energetic, informational, or both, can be released through various strategies. These strategies, implemented by the listener, may operate at different levels of awareness and involve more or less LE. In fact, we constantly and automatically inhibit irrelevant sounds in our environment. These strategies can be automatic or deliberate and effortful, depending on how demanding the situation is. In everyday environments, we often suppress irrelevant auditory information without noticing. It is only when the auditory scene becomes challenging and effort is required that we become aware of this release process (Pichora-Fuller et al., 2016).

Several cues, characterized by the sound sources, can be used by the listener to release from masking in order to increase SI and decrease LE (see Box 4.2).

Masking release strategies

- Location of the talkers, spatial separation between the target and the maskers (Rennies et al., 2019)
- Pitch difference, such as variation in fundamental frequency (F0, see Box 4.2) of the talker(s) voice(s)
- Talker familiarity, a familiar voice is easily segregated from other voices in cocktail party (Johnsrude et al., 2013)
- Temporal modulation and roughness
- Talkers gender (Brungart, 2001a)
- Languages of the talkers (see Section 4.2) (Cooke et al., 2008; Van Engen and Bradlow, 2007)
- Loudness differences

For more details, Culling and Stone (2017) provides explanations of the different types of EM release, including spatial separation, envelope, fluctuations, and differences in F0.

Fundamental Frequency

Fundamental Frequency (F0) of a sound is the lowest frequency of a complex sound wave (for example speech). It determines the periodicity in the acoustic waveform of voice and thus the perceived pitch of the sound. The human voice F0 is generally located between 70 and 280 Hz.

Speech in speech - TMR

In multi-talker condition studies, also called speech-in-speech (SIS), it is common to get the participants to pay attention to one talker in particular (with different types of cues depending on the corpus and the focus of the study). The difference in sound level between the target talker and the masker(s) is called the target-to-masker (TMR) ratio (in dB, see Figure 1.9). This ratio is positive when the target has a higher sound level than the masker, negative when the target has a lower sound level than the masker, and neutral when the target and the masker have an equal sound level.

In a single masker SIS condition, intelligibility is influenced, among other factors, by the similarity between the target and masker voices. The closer the voices in their acoustic characteristics, the more challenging the segregation of the target speech, resulting in poorer performance (Brungart, 2001b). SI also improves when the target and masker differ in gender. The most detrimental condition occurs when the target and masker are produced by the same talker. This is consistent with the concept of IM: the more acoustically and perceptually similar the competing voices, the harder the separation, thus increasing LE. In SIS, it is complex to dissociate

EM from IM. Additional talkers not only introduce IM through the concurrent and competing speech information but also contribute to EM due to the added acoustic energy of their voices. However, using SIS is a common approach to investigate IM specifically (Brungart, 2001b).

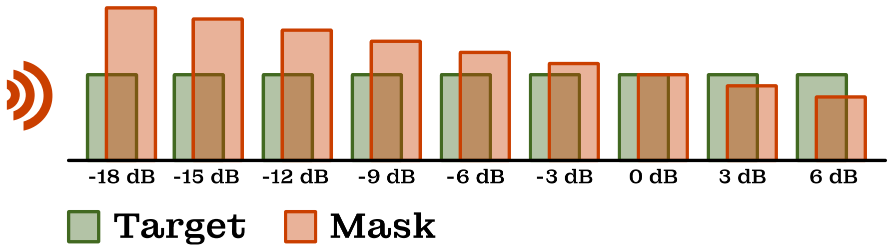


Figure 1.9: Visual representation of TMR and SNR with the Target (talker) in green and the Mask (talker or noise) in orange.

Speech in Noise - SNR

For the study of the impact of EM on SI and LE, speech-in-noise (SIN) scenarios are often used. In these cases, noise is added to the target talker stream. This noise can be of diverse origins, such as white noise or spectrum-shaped noise, or stationary noise (considered as pure EM). The sound level difference between the target talker and the noise is called signal-to-noise ratio (SNR).

> **Nomenclature Clarification**
>
> In this project, we refer to the ratio in SIN situations as the signal-to-noise ratio (SNR), and in SIS situations as the target-to-masker ratio (TMR). The concept is the same (see Figure 1.9): the difference in sound level between the target talker and a masker (noise or another talker). We chose this nomenclature for clarity.

Spatial release of masking

In real-life situations, the spatial location of talkers is an important factor contributing to masking release. Listeners can use spatial cues to segregate the target sound from competing sounds (Andéol et al., 2017; Middlebrooks et al., 2017). These cues depend not on the sound itself, but rather on where the sound source is located within the auditory scene. Therefore, spatial localization has a direct impact on SI and LE (Andéol et al., 2017; Darwin, 2008). Moreover, when sound sources remain in fixed positions, streams are easier to segregate, allowing the listeners to rely on spatial cues for effective masking release (Brungart and Simpson, 2007).

Spatial localization of sound

The human auditory system, with two ears separated by the head, enables spatial hearing through interaural differences. A sound will reach one ear slightly earlier (interaural time difference) or louder (interaural level difference) than the other, depending on the source's position. These interaural cues are crucial for localizing sounds and play a key role in auditory scene analysis (Middlebrooks, 2015).

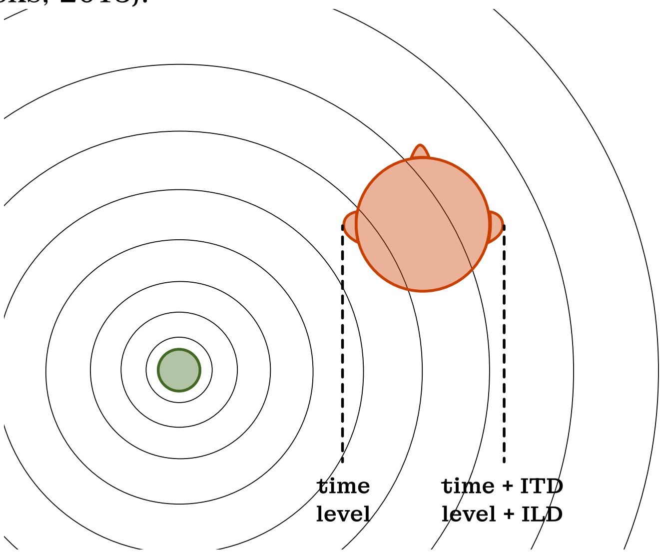


Figure 1.10: Example of source localization with in green: the auditory source, in orange: the listener, ITD: interaural time difference, ILD: interaural level difference

Many studies investigating EM and IM incorporate spatial cues in their experimental designs. For example, dichotic listening tasks or spatially separated sound sources are frequently used to examine the effects of stream localization on SI or LE.

Although spatial localization of sound is an essential and widely studied mechanism in natural listening, it is not the focus of the present work. For in-depth reviews on spatial stream segregation and spatial hearing mechanisms, please refer to Darwin, 2008; Middlebrooks, 2017.

Listeners' knowledge of the language and voice familiarity

In a complex auditory scene, the listener's familiarity with the language spoken and the identity of the talkers can influence both SI and LE. Talkers (target or masker) may not all speak the same language or may use an unknown or non-native language for the listener. The similarity between target and masker voices plays a role in SI. As noted by Rhebergen et al. (2005), the more similar the voices, the greater the confusion or the distraction, leading to poorer comprehension.

In SIN conditions, native listeners tend to perform better when the target speech is in their native language than when it is in a non-native one (Lecumberri and Cooke, 2006; Lecumberri et al., 2010; Rogers et al., 2006). Non-native listeners, however, suffer from increased EM in SIN conditions (Cooke et al., 2008), as their ability to segregate speech is more affected by the noise.

In SIS conditions, several language configurations are possible. The target and masker may both speak the listener's native language, both may speak a non-native or unknown language, or each may use a different language. Indeed, when the target talker speaks the listener's native language and the masker speaks a non-native one, it becomes easier for the listener to segregate the speech streams, thus improving SI. Additionally, masking in a familiar language can increase IM compared to when the masker speaks a non-native language (Lecumberri and Cooke, 2006; Rhebergen et al., 2005; Van Engen and Bradlow, 2007).

Beyond language, voice familiarity also contributes to improved speech comprehension. Listeners understand speech more easily when the target voice is familiar, particularly in SIS scenarios (Johnsrude et al., 2013). This effect occurs both when the familiar voice is the target or the masker, suggesting that familiarity helps the formation of auditory objects and the segregation of speech sound sources.

> **Conceptual Clarification**
>
> In this project, we refer to adverse conditions as situations in which listening difficulty is expected to increase, whereas favorable conditions correspond to situations in which difficulty supposedly decreases. Mattys et al. (2012) defines adverse condition as "any factor leading to a decrease in speech intelligibility on a given task relative to the level of intelligibility when the same task is performed in optimal listening situations".

5 Why studying listening effort

Hearing loss affects over 5% of the world population. According to the World Report on Audition (Organization, 2021), nearly 2.5 billion people will experience some form of hearing loss by 2050, meaning that one in ten individuals will be affected during their lifetime. These numbers underline the growing importance of addressing hearing-related issues and their broader consequences on health and quality of life.

In daily life, oral communication plays a central role in social interactions. Difficulty understanding speech, especially in complex auditory environments, can make it harder to engage in conversations. Over time, this can lead to social withdrawal, increased isolation, or even mental health difficulties (Pichora-Fuller et al., 2016).

Hearing difficulties are often thought to result solely from frequency deficit and can be quantified with audiometry. However, a significant and often overlooked issue involves people who report auditory difficulties despite having normal hearing thresholds on tonal audiometric tests. This issue, known as hidden hearing loss (HHL), is difficult to detect with conventional tools but can significantly affect everyday communication and well-being. People affected by HHL often have difficulties following conversations in groups or in noisy environments, increased sound sensitivity, tinnitus, or hyperacusis.

HHLs are related to a general decrease of SI and an increase in LE as soon as the auditory scene becomes complex. A constant high-exertion in LE to compensate for these challenges can result in chronic stress and a deterioration in overall well-being

(Mattys et al., 2012; Pichora-Fuller et al., 2016).

Furthermore, although HHL is more frequently observed in older adults, it increasingly affects younger populations, particularly those exposed to loud noise environments (headphones, parties; Zheng and Guan, 2018). Also, people working or living with frequent exposure to noise, such as teachers, construction workers, military personnel, plane pilots, or employees in open-plan offices, are particularly at risk. Studies show that 5 to 12% of people with normal audiometric thresholds still report difficulty understanding speech, especially in challenging listening situations (Tremblay et al., 2015). Known contributing factors include noise exposure, use of ototoxic medication, viral infections, and certain genetic mutations (Zheng and Guan, 2018).

Hearing loss is often associated with damage or degeneration of hair cells in the cochlea. However, a recent hypothesis for the cause of HHLs is cochlear synaptopathy leading to damage to synapses located between the inner hair cells of the cochlea and the auditory nerve fibers (Liberman and Kujawa, 2017; Zheng and Guan, 2018). These synapses are crucial for transmitting acoustic information to the brain (see Chapter 2). Even in the absence of hair cell damage, the degradation of these connections can happen and disrupt auditory processing and remain hidden from audiometric tests (Liberman and Kujawa, 2017; Liberman et al., 2016; Organization, 2021). Nevertheless, at low intensity of cochlear damage, a small increase in sound level can compensate, making it at first difficult to distinguish (Liberman and Kujawa, 2017). For a detailed explanation of mechanisms of the cochlear synaptopathy, please refer to Liberman and Kujawa (2017).

Thus, since most cases of HHLs are acquired rather than congenital, preventive strategies are both important and possible. Despite the clinical relevance, standard diagnostic tools are still missing. Zheng and Guan (2018) emphasize the need for a diagnostic test battery that could improve accuracy and help differentiate between types of hearing impairment. Some studies on neuroplasticity in humans and animals suggest that auditory training may help improve hearing perception associated with HHL (Whitton et al., 2014).

Studying LE is thus essential in this context. A better understanding of mechanisms and reliable brain measures of LE would support early diagnosis, better rehabilitation, and improve clinical care (McGarrigle et al., 2014).

6 Conclusion on the chapter

The auditory scene is often complex, requiring listeners to deploy strategies to release both energetic and informational masking. The associated listening effort, closely linked to speech intelligibility, varies between individuals and remains difficult to objectively quantify. Although several physiological and behavioral methods exist to measure listening effort under experimental conditions, none fully capture the related cognitive mechanisms, and its definition is still discussed in the literature.

Because listening effort and speech intelligibility are cognitive constructs, a deeper understanding requires considering the neural mechanisms that support them. Therefore, in the next chapter, we introduce the brain pathways underlying speech comprehension processing, with a particular focus on how they relate to listening effort.

Summary of the chapter

- The auditory scene is often complex, and listeners experience listening effort differently.
- Listening effort is a multidimensional concept that varies between subjects.
- Physiological measures of LE exist, notably electroencephalography or pupillometry.
- Self-assessment measures allow a subjective measure of LE.
- Research on LE is essential for a better understanding of the concept, allowing for better diagnosis procedures and clinical care.

2

The Auditory Brain

Description of the Chapter

Once the auditory scene and listening effort are described, the question of how the brain is involved arises. Brain research explores the neural basis of human cognition, investigating how different brain regions and networks support a wide range of functions, including speech processing. In this chapter, we introduce the brain pathways related to speech understanding and, more precisely, to listening effort. Additionally, to prepare for the later scientific contribution using electroencephalography as a physiological measure of listening effort, the technique is briefly explained, along with different analysis methods.

1 Speech and audition perception

Why is speech a specific sound?

The human auditory system continuously receives complex sound waves that represent the entire auditory scene surrounding the listener. All incoming sounds arrive as a single, combined signal in the auditory system. Furthermore, speech is a specific type of sound.
Speech stands out due to its unique combination of acoustic, linguistic, and communicative characteristics, allowing us to distinguish and interpret it differently from other types of sounds (Bronkhorst, 2015). For a detailed review on how the physics of speech, including frequency selectivity, timbre perception, pitch perception, and temporal analysis, is specific in hearing, please refer to B. C. Moore (2008).
When processing speech, the brain recruits specific neural pathways that differ from those used for general sound perception. While many aspects of speech processing overlap with other sounds, there are unique brain mechanisms specialized for speech (Bronkhorst, 2015). Also, top-down processes enable the brain to automatically extract meaning from speech signals (Davis and Johnsrude, 2007, see Box 1.5).
Speech is unique due to the distinctive way the brain interacts with it (D. R. Moore, 2000).

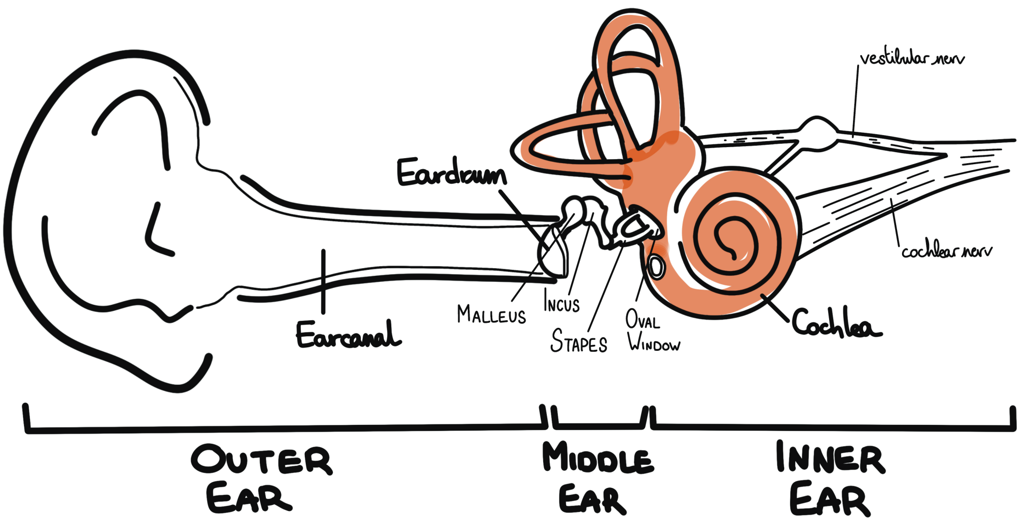


Figure 2.1: Human ear anatomy with the outer, middle, and inner ear.

1.1 The Ear Sound perception

Periphery: Outer, middle, and inner ear

The human ear consists of three main parts: the outer ear, the middle ear, and the inner ear (see Figure 2.1). As sound enters the outer ear, it travels through the auditory canal, where it is amplified until it reaches the tympanic membrane. The tympanic membrane is in direct contact with the malleus, one of the three ossicles of the middle ear, diffusing the vibration to the incus and stapes. The stapes transmits the sound to the cochlea in the inner ear through the oval window.

The cochlea is composed of three distinct chambers (see Figure 2.3): the scala vestibuli, (upper chamber), the scala media (cochlear duct), and the scala tympani (lower chamber). The scala vestibuli and scala tympani are filled with perilymph, while the scala media contains endolymph, a potassium-rich fluid crucial for hair cell activation.

Cochlea

The cochlea is the structure in which sound waves are transformed into electrical signals. This snail-shaped organ plays a crucial role in hearing, and damage to it is often irreversible, making it a common source of hearing disorders. The cochlea is organized in a tonotopic manner, meaning that the sound frequencies are spatially coded along the cochlea, from high frequencies to low frequencies. Each cochlea contains approximately 15,000 hair cells and 40,000 nerve fibers. These cells lack regenerative capacity; thus, once damaged or lost, their function cannot be restored, resulting in irreversible auditory impairment.

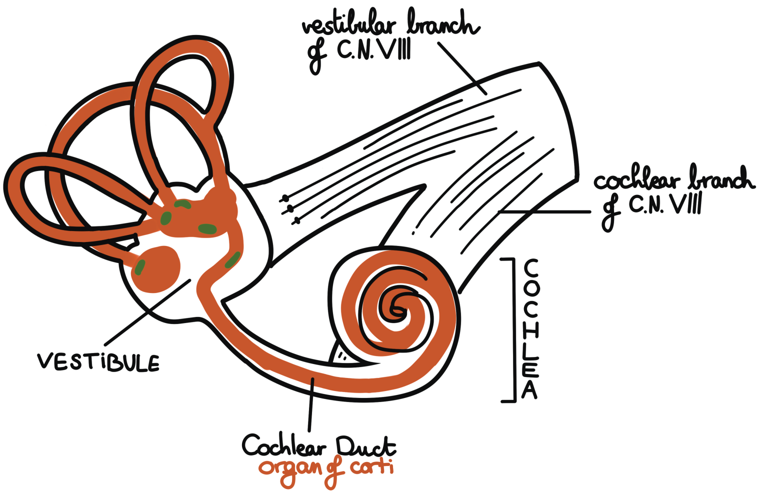


Figure 2.2: Inner ear and cranial nerve VIII.

A movement of endolymph in the scala vestibuli is initiated by vibration of the ossicles via the oval window and travels along the cochlea. It leads to movements of hair cells located in the organ of Corti, conducting to the depolarization of hair cells. These hair cells are responsible for frequency perception (Box 1.1). They convert mechanical movement into neural signals, which are then sent to the brain stem and brain through the auditory nerve.

Cochlea to Cortex

The hair cell activation creates a signal transmitted through the cochlear nerve (part of cranial nerve VIII). This nerve exits the cochlea in the direction of the brainstem,

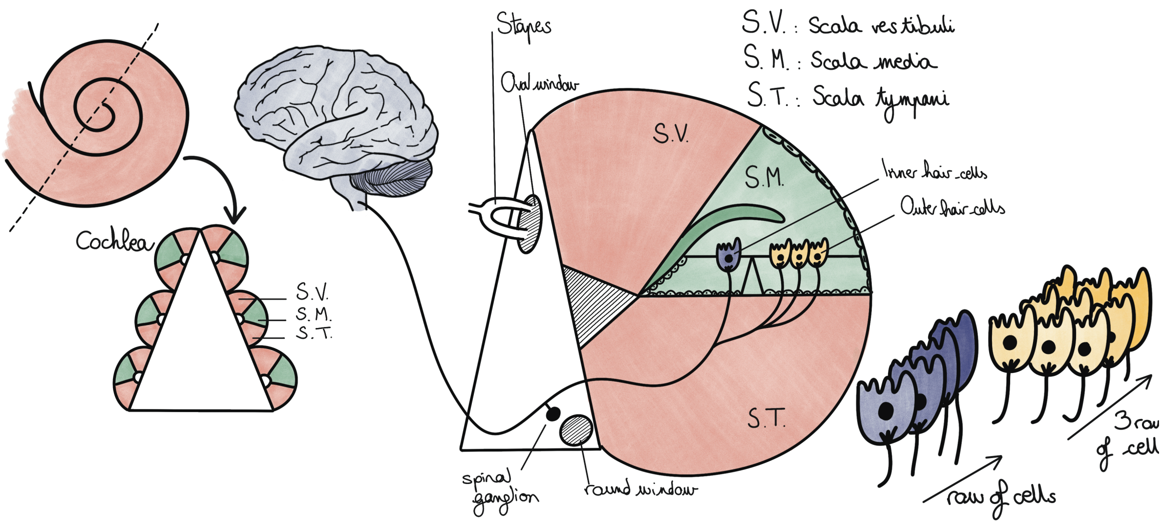


Figure 2.3: Cochlea section representing the scala vestibule, scala media, and scala tympani.

projecting to the cochlear nuclei. These nuclei are the first processing step in the brain's processing of sound through the primary auditory pathway (see Figure 2.7). The signal is sent both ipsilaterally and contralaterally to the superior olivary complex, which plays a key role in sound localization. From there, the sound moves through the lateral lemniscus, reaching the inferior colliculus and then the medial geniculate body in the thalamus (Figure 2.4).

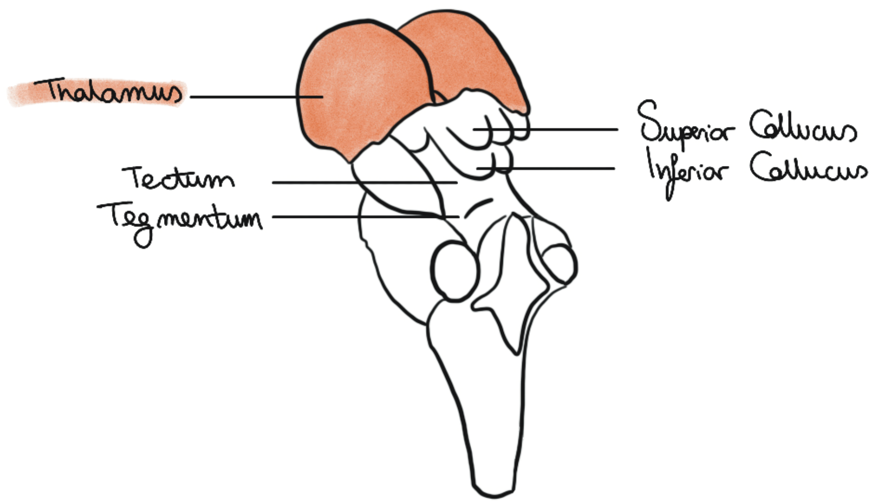


Figure 2.4: Brain stem with Thalamus.

Finally, the signal reaches the primary auditory cortex (Box 1.1) in the superior temporal gyrus of the brain. This is where the sound signal will be translated in order of frequencies and redirected to different cortical areas for processing and interpretation.

Inferior frontal gyrus

The inferior frontal gyrus (IFG), situated in the frontal cortex between the inferior frontal sulcus and the superior temporal sulcus, is anatomically divided into three subregions: the pars orbitalis, pars triangularis, and pars opercularis. The left IFG, often referred to as Broca's area, is critically involved in both language perception and production. The left IFG includes both language-selective and domain-general areas (Fedorenko et al., 2012). Beyond its linguistic functions, the IFG also contributes to executive processes, reflecting its integration within broader cognitive control networks.

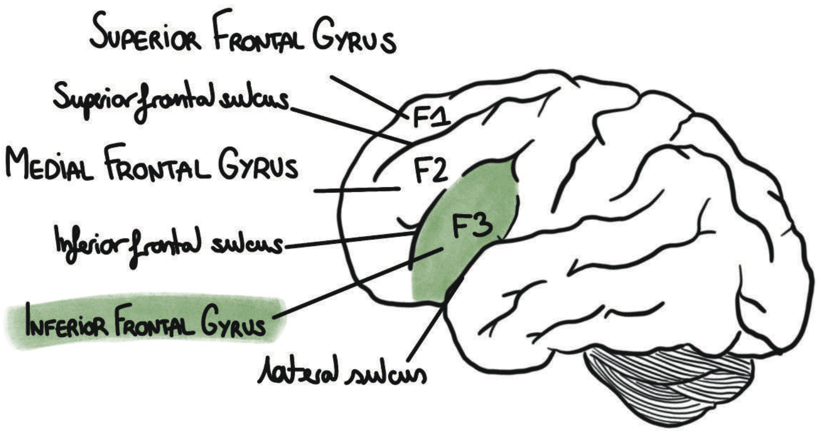


Figure 2.5: Inferior frontal gyrus

Auditory Cortex

The auditory cortex, located in the superior temporal gyrus, is hierarchically organized: the core (including primary auditory cortex) processes simple acoustic features; the belt region processes more complex spectrotemporal patterns; and the parabelt integrates information with other modalities and cognitive processes (Chandrasekaran et al., 2022).
Like the cochlea, the primary auditory cortex is organized in a tonotopic manner, meaning that it spatially maps the cochlear frequencies (Purves et al., 2001) as shown in the following Figure.

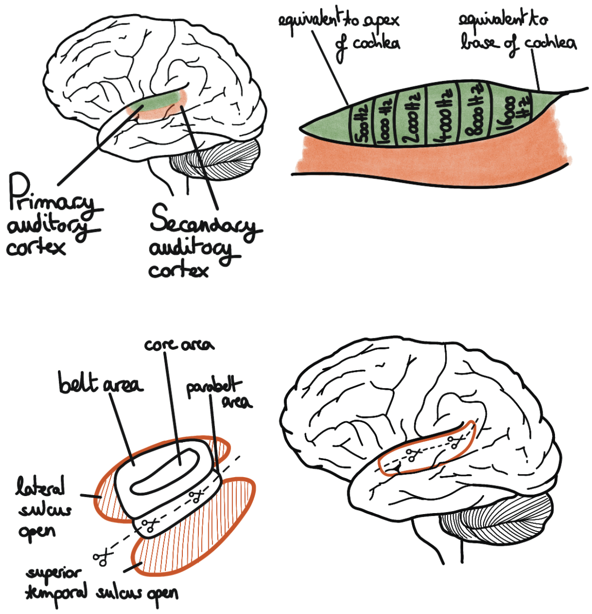


Figure 2.6: Primary and secondary auditory cortices

1.2 After the Ear: the Brain Speech Perception

Cortical Pathways

As explained above (Box 1), speech is a specific sound. Therefore, the neural pathways involved in speech perception exhibit some specificity compared to those processing non-speech sounds. Nevertheless, it is important to note that speech and non-speech sound perception interact and share overlapping brain pathways (Holt et al., 2022). In contrast, inner-ear activity does not differ according to whether the sound contains speech or not.

Broca and Wernicke were among the first to propose the localization of language functions in the brain. Now identified as Wernicke's area (superior temporal gyrus) as essential for language comprehension and Broca's area (inferior frontal gyrus) as crucial for language production. While these areas were initially identified based on the effects of brain lesions, later evidence showed that damage to these regions does not always result in language deficits, suggesting a more distributed organization.

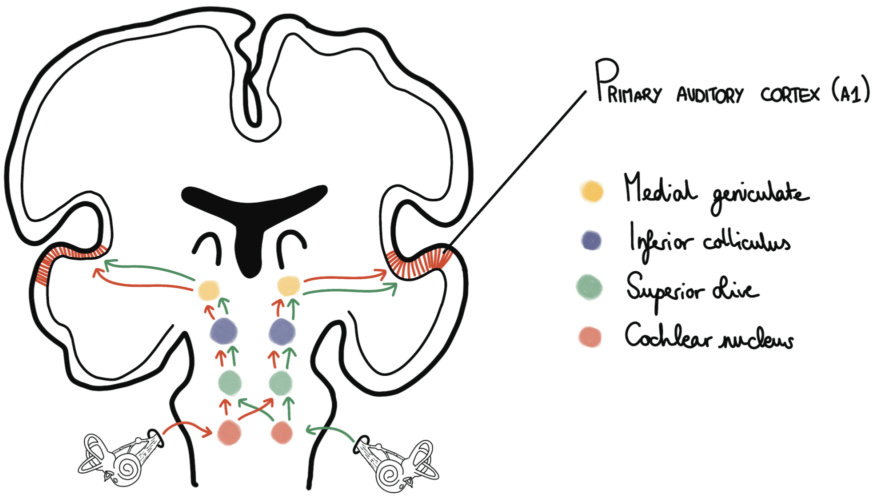


Figure 2.7: Primary auditory pathways. Coronal section of the brain and brain stem.

Modern neuroimaging techniques such as fMRI, PET, MEG, fNIRS, or EEG have enabled the exploration of the neural pathways underlying speech perception. Despite significant progress and years of brain research, a fully detailed and unified model of the neural organization supporting speech perception has not been established yet (Adank, 2012; "Brain Mechanisms of Auditory Scene Analysis", 2020; Hickok and Poeppel, 2007).

In the cortex (Figure 2.8), speech is processed within a bilateral core language network, supplemented by additional areas depending on task demands and processing level (Hertrich et al., 2020; Peelle, 2018). Cortical activity is responsible for the transformation of lower-level auditory features into higher-level representations, which are fundamental to understanding speech (Chandrasekaran et al., 2022).

Commonly identified cortical regions implicated in speech perception include the auditory cortex, the inferior frontal gyrus (IFG), Broca's area, the superior temporal gyrus (STG), Wernicke's area, the superior temporal sulcus (STS), and the middle temporal gyrus (MTG) (Friederici, 2011; Friederici and Gierhan, 2013).

However, the pathways connecting these areas are task dependent (Hickok and Poeppel, 2007). Similarly to vision brain networks, a common model describes two major processing streams, a dorsal and a ventral pathway (Friederici and Gierhan, 2013; Hickok and Poeppel, 2007.

Dorsal pathway The dorsal pathway is mainly associated with speech perception (Friederici and Gierhan, 2013) and is involved in auditory-motor integration (Chang et al., 2015; Hickok and Poeppel, 2004). It sends projections from the STG towards the parietal lobe and then to frontal regions such as the IFG (Box 2.5; Friederici, 2011; Hickok and Poeppel, 2007).

Ventral pathway The ventral pathway is responsible for speech recognition, particularly semantic processing of speech, by mapping sound onto meaning (Friederici

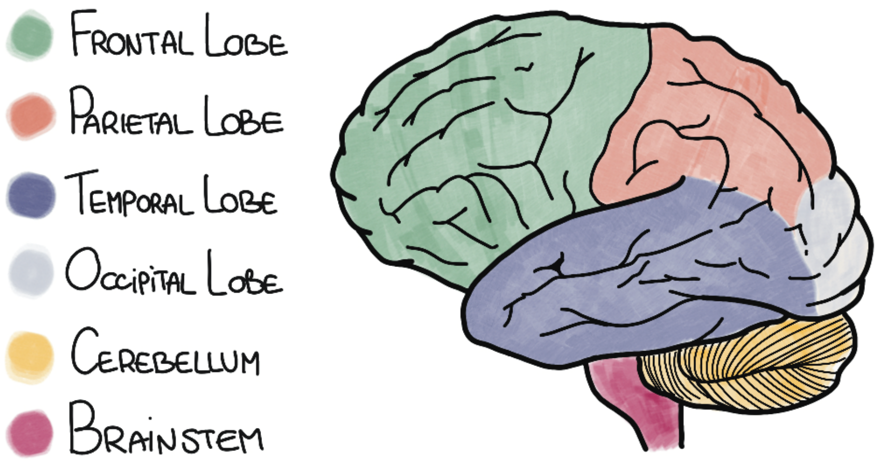


Figure 2.8: Brain Lobes. Cortical areas of the brain.

and Gierhan, 2013; Hickok and Poeppel, 2007) and projecting to the anterior and middle temporal lobes (Chang et al., 2015).

Although the dorsal-ventral model is widely adopted, it likely captures only part of the complex network underlying speech perception.

The precise correlates of language perception in the brain remain difficult to define. While many studies aim to describe a consistent language network, findings suggest that language processing might rely on multiple overlapping and dynamic networks. These may vary depending on the specific process, the listening condition, task demands, and broader context. It might be that language is related to multiple networks, depending on the process, task, context, or auditory situation. For a more detailed overview of cortical language networks and their roles in speech perception, please refer to Hertrich et al. (2020) or Arnal et al. (2016).

1.3 The speech network in complex auditory situations

During degraded speech perception, brain activity changes to compensate for reduced intelligibility (Peelle, 2022). These changes may help identify neural markers of listening effort related to increased difficulty and decreased speech intelligibility.

Functional neuroimaging studies consistently show increased brain activity in complex auditory situations, in the speech-related areas especially (Alain et al., 2018). Activity in core speech regions, such as the STG and the STS, tends to increase linearly as speech becomes less intelligible. In contrast, non-linear activation patterns have been observed in other regions, such as the IFG, particularly the left IFG (Alain et al., 2018; Binder et al., 2004; Lanzilotti et al., 2022; Lawrence et al., 2018; Peelle, 2022; Wild et al., 2012; Zekveld et al., 2006). For instance,Wild et al. (2012) report that left IFG activation peaks at intermediate level of intelligibility, forming an inverted U-shaped response, also observed in pupillometry measures in complex auditory conditions (Zekveld et al., 2014). The authors suggest that the

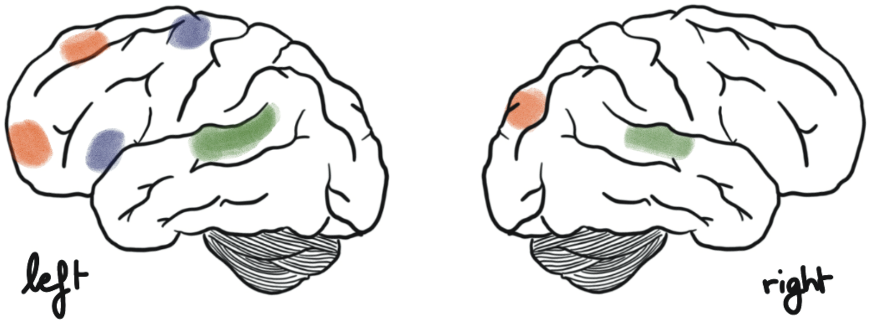


Figure 2.9: Cortical areas associated with different aspects of speech perception in complex auditory situations. Orange: brain areas more activated by SIN than SIS. Green: brain areas more activated by SIS than SIN. Blue: brain areas which increase in activity as the SNR of SIN decreases. - adapted from Scott and McGettigan, 2013.

left IFG activation pattern could be a neural marker of listening effort. In a very detailed review on neural processing of masked speech, Scott and McGettigan (2013) describe different activation patterns in SIS and SIN situations, illustrating that some areas in particular, such as the dorsolateral temporal lobes, are differently activated depending on the masking type (see Figure 2.9).

In more adverse listening conditions, additional brain regions are recruited, including the cingulo-opercular network (CON), which is commonly associated with domain-general cognitive control and task engagement (Peelle, 2022).

Most fMRI or fNIRS studies focus on conditions in which target speech is masked by background noise (SIN). Some authors (Evans et al., 2016) suggest that SIS, with two concurrent streams, should show the same activation pattern as the SIN scenario. However, Lanzilotti et al., 2022 showed that, in opposition to SIN studies and previous results, the left IFG activity decreased in adverse conditions when individuals are trained. This relationship between left IFG activity and listening in effortful situations could suggest the left IFG as a potential neural index of LE, and thus as a potential objective marker of it.

2 Electroencephalography

Several tools are available to measure brain activity, each capturing different aspects of neural function. Electroencephalography (EEG) specifically provides non-invasive recordings of electrical fluctuations generated by neuronal activity. First described by Hans Berger in 1929 (Berger, 1929), EEG has become a widely used technique in applied and clinical neuroscience, as well as for clinical research and diagnosis.

Today, EEG is a standard method in cognitive neuroscience, mainly due to its high temporal resolution (on the order of milliseconds) and relatively low cost compared to imaging techniques such as fMRI, PET-scan or MEG. Its ease of installation, portability, and non-invasiveness make it particularly well-suited for clinical applications and experimental protocols that require rapid setup and flexible use in laboratory or more ecological environments.

As illustrated in Figure 2.10, EEG provides good temporal resolution, as it directly captures neuronal electrical activity from scalp electrodes, but suffers from limited spatial resolution – at least from the raw signal – due to the volume conduction and filtering effects of the skull and other tissues.

The number and configuration of electrodes play a critical role in determining the localization of current source generators (or electrical dipoles) and the types of analyses that can be performed. Standardized systems, such as the 10-20 or 10-10 international systems, facilitate reproducibility and compatibility between studies. Given its long-standing use and broad adoption in various disciplines, EEG benefits from a rich ecosystem of analysis tools, methodological guidelines, and dense documentation, supporting a wide variety of research applications. Recent developments in signal processing and machine learning further enhance the utility of EEG, with significant impact expected in the coming years (Mushtaq et al., 2024).

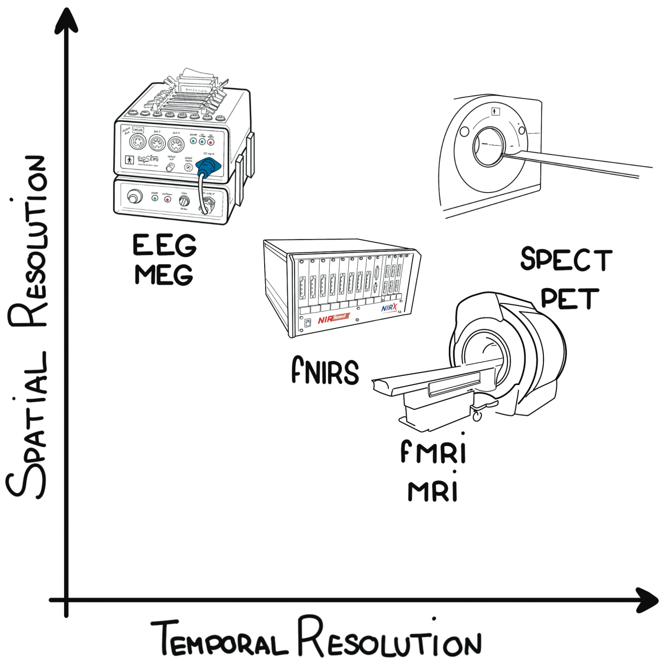


Figure 2.10: Brain Imaging methods. Temporal and spatial resolutions of different brain imaging types. EEG: electroencephalography, MEG: magnetoencephalography, fNIRS: functional near-infrared spectroscopy, fMRI: functional magnetic resonance imaging, SPECT: single-photon emission computed tomography.

2.1 Physiological aspects

In the brain, neurons exhibit a wide variety of sizes, shapes, and purposes. A typical neuron, such as a pyramidal neuron (see Figure 2.11), consists of a soma, axons, dendrites, and synapses. Neuronal communication is primarily driven by the movement of sodium (Na+) and potassium (K+) ions across the axon membrane, producing an action potential that propagates along the axon (see Figure 2.12). At rest, neurons maintain a membrane potential of approximately -70 mV, due to ion concentration gradients outside and inside the axon membrane. When neurotransmitters bind to the membrane receptors, a cascade of events is initiated that alter the local ion concentrations. If the change in membrane potential reaches a certain threshold, voltage-gated ion channels open, resulting in a rapid depolarization: the neuron fires, and an action potential travels down the axon to the synapses.

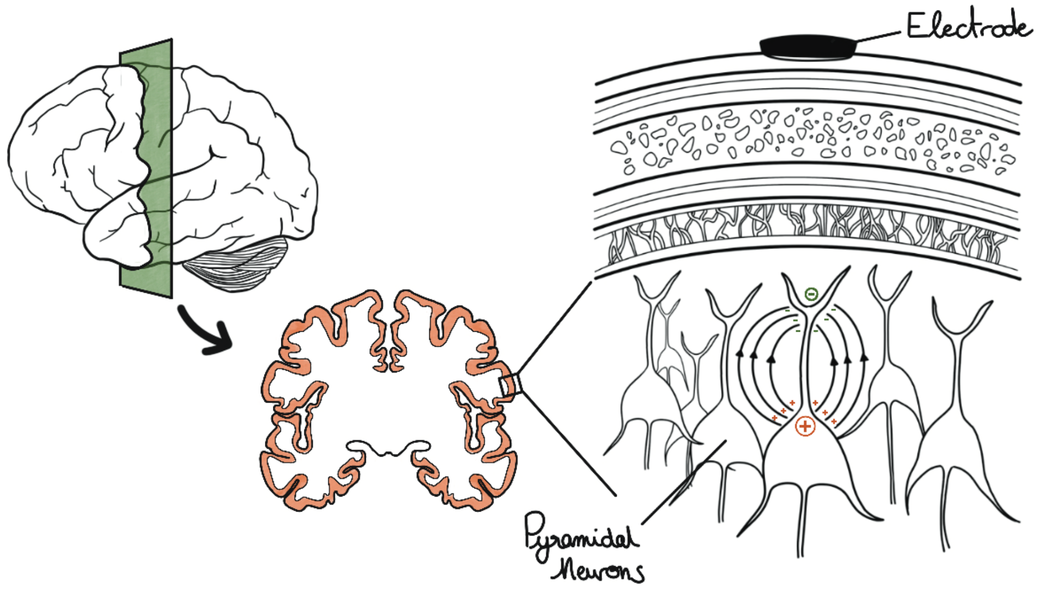


Figure 2.11: Coronal section of the cortex and representation of cortical pyramidal neurons with the electrical field they create.

EEG does not directly measure action potentials, but rather the summed post-synaptic potential across large populations of neurons. These excitatory or inhibitory potentials generate measurable electric dipoles in the brain.

Post synaptic potentials

In the synaptic area, neurotransmission leads to chemo-dependent or voltage-dependent gated channels opening, which permits the generation of an electrical potential in the post-synaptic area. These potentials will stimulate or inhibit the neuron (see Figure 2.12).

- Excitatory post-synaptic potential: Sodium ions (positively charged) (Na+) flow into the cell, making the intracellular area more positive and the extracellular area more negative near the synapse. This shift causes the distant extracellular space to become relatively more positive, forming a dipole. If the Excitatory postsynaptic potential (EPSP), or a sum of EPSPs, generates a strong enough potential, an action potential will be generated.

- Inhibitory post-synaptic potential: in opposition, the post-synaptic area hyperpolarized, reducing the likelihood of firing and creating an inverse dipole configuration.

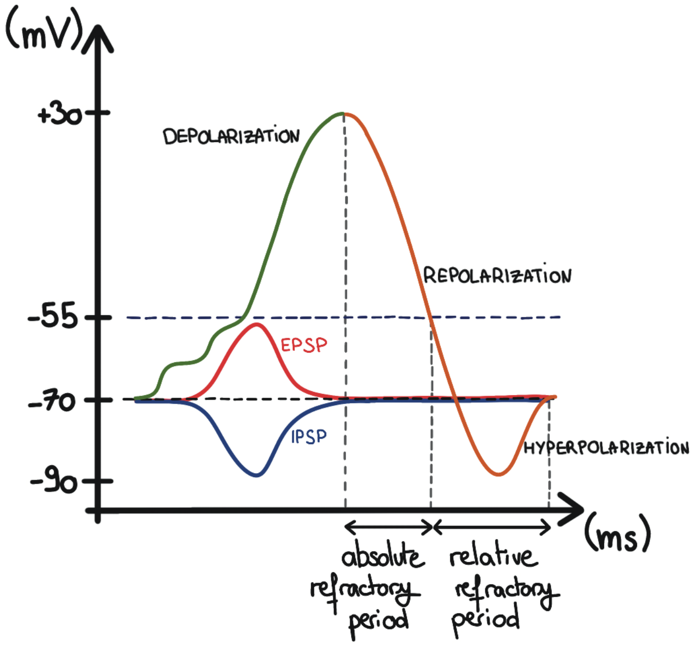


Figure 2.12: Typical action potential waveform of a neuron showing membrane voltage (mV) over time (ms). EPSP: Excitatory postsynaptic potential. IPSP: Inhibitory postsynaptic potential.

The cerebral cortex is folded into gyri and structured in well-defined layers. Among cortical neurons, pyramidal neurons in layers III and V are of particular importance for EEG signal recording. These neurons are perpendicularly aligned to the cortical surface and oriented in parallel, enabling the spatial summation of local field potentials generated by synchronized local field potentials. This synchronized activity allows for the recording of summed electric fields strong enough to be recorded at the scalp. Active electrodes displayed on the scalp detect these electrical fluctuations, which can then be amplified, processed, and analyzed for interpretation.

The electroencephalogram reflects the activity of cortical neurons, while deeper brain structures contribute indirectly, as discussed in the Section 2.4.

The EEG recording is performed by placing a predefined number of electrodes on the scalp, following the international 10-20 system (see Box 2.1). This system facilitates the standardization of electrode locations at regular intervals on the scalp. The number of electrodes has a large impact on the resolution of the recording. The more electrodes are displayed, the higher the spatial resolution.

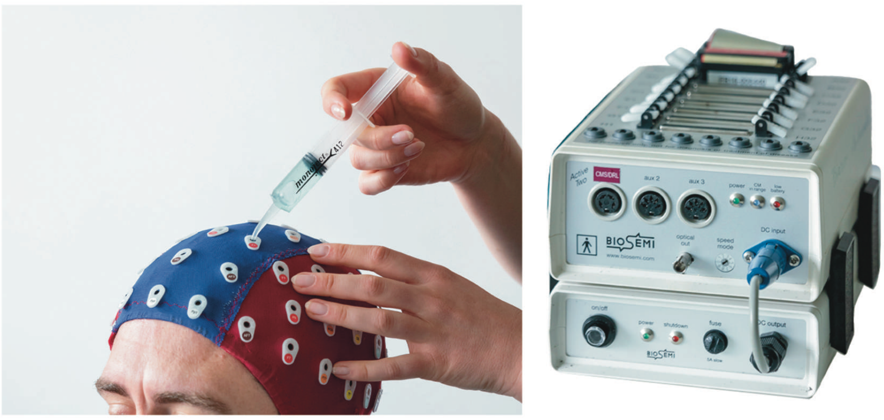


Figure 2.13: EEG system. Left: installation of a 64-electrode EEG cap using conductive gel. Right: Biosemi Active II amplifier.

10 20 System

The 10-20 system, or international 10-20 system, is widely used in electroencephalography. This system is a standardized method for the electrodes' placement on the scalp, which was proposed by Jasper and Andrew (1938). The system provides precise electrode positions at 10% or 20% of the distance between the inion and the nasion, as well as between the preauricular points. Each electrode is labeled according to its position (F: frontal, C: central, T: temporal, P: parietal, z: midline), with odd numbers referring to electrodes on the left hemisphere and even numbers to those on the right hemisphere. In the current project, the use of a 64-electrode EEG device implies using the 10-10 system (10% distance in each direction).

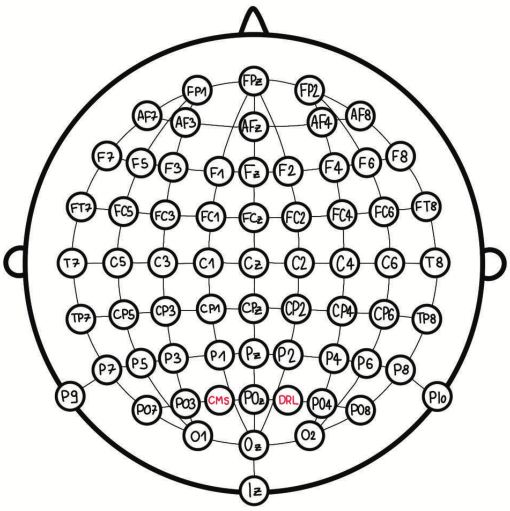


Figure 2.14: 10 20 System

In addition to placement, different types of electrodes can be used. They can be

”dry” (direct contact with the scalp) or ”wet” (with a conductive gel), pre-amplified or not pre-amplified.

2.2 Brain Oscillations

The EEG signal is characterized by a continuous voltage recorded at the scalp that varies in amplitude and frequency over time. In healthy conditions, the human brain oscillates at frequencies between 0.05 and 500 Hz (Buzsáki and Draguhn, 2004) with amplitudes between 20 to 100 μV. The brain oscillates constantly under different rhythms, which are usually classified according to their frequency. Although often presented as distinct bands, actual EEG signals are a mixture of overlapping rhythms, reflecting the simultaneous activity of multiple neural processes.

Brain oscillations are commonly divided into five main frequency bands (Figure 2.15): delta (<4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (13-30 Hz) and gamma (¿30 Hz) bands. These bands are associated with different functional states of the brain and are often related to different sleep and wakefulness stages. High-frequency oscillations tend to arise from more localized neuronal populations, while lower ones recruit larger brain networks (Buzsáki and Draguhn, 2004). Because amplitudes decrease as frequency increases, high-frequency oscillations were only identified later in EEG research, and it is now known that oscillations to 600 Hz can be observed (Curio et al., 1994; C. S. Herrmann and Demiralp, 2005). As given frequency bands can occur at different brain states and anatomical regions, Hari and Puce (2023) suggest that both frequency and anatomical location should be reported when describing brain rhythms.

delta δ (<4 Hz) First described by Walter (1936), delta oscillations are most prominent during deep sleep and are thought to originate from the medial frontal cortical regions (Knyazev, 2012).

theta θ (4-8 Hz) Also described by Walter (1936), theta oscillations are initiated in the hippocampus (Buzsáki and Draguhn, 2004; Lu et al., 2020) and are linked to cognitive control, working memory, and spatial navigation. They are also involved in memory processing for visual (Moreau et al., 2020) and auditory (Teng and Poeppel, 2020) information. Frontal midline theta (Fmθ, 4-8 Hz) generated by medial frontal sources (Ishii et al., 1999) is associated with attention, memory, and task difficulty (Onton et al., 2005; Wisniewski et al., 2015) and is also used as a neural index of LE (Mohammadi et al., 2024; Wisniewski et al., 2015).

alpha α (8-12 Hz) Observed by Berger (1929), alpha rhythms were the first described oscillations and are among the most studied. They are especially visible during relaxed, eye-closed resting states, particularly over occipital regions. Alpha activity decreases with attention and alertness and is linked to thalamo-cortical synchronization (Beppi et al., 2021). These oscillations are related to multiple sensory domains, including audition (Weisz et al., 2011). Changes in alpha power have also been associated with listening effort (Wisniewski et al., 2017). Subtypes include mu rhythms (9-11 Hz), related to motor activity, and tau rhythms (8-13 Hz), which are associated with auditory processes (see Chapter 1, Section 3.3). Tau rhythms are not easily visible in raw EEG data and are sometimes better captured

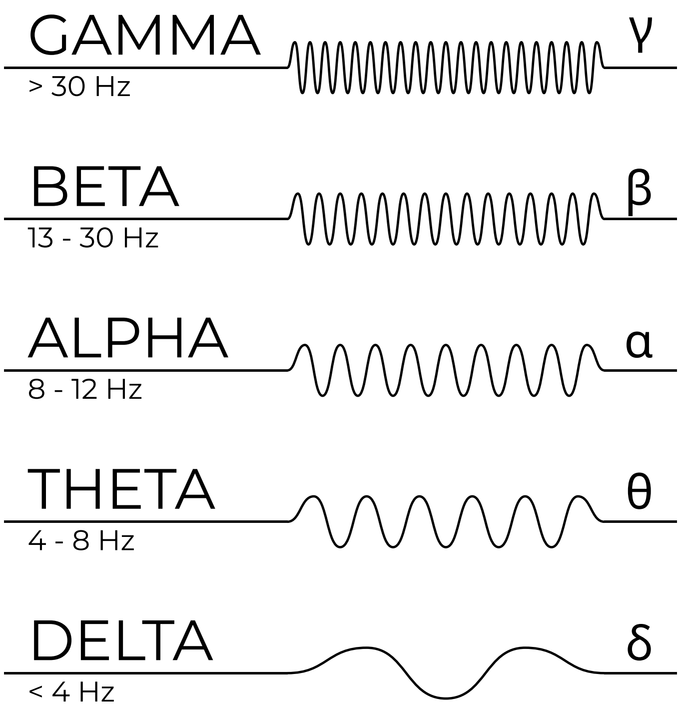


Figure 2.15: Brain oscillations across different frequency bands: delta (<4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (> Hz).

with MEG, although EEG-based approaches have also been developed (Wisniewski et al., 2024).

Tau rhythms

Tau rhythms are a subtype of alpha oscillations, typically ranging from 8 to 13 Hz, and are particularly associated with auditory processing and speech listening, especially in adverse listening conditions (Lehtelä et al., 1997). Tau activity has been localized to the parieto-temporal regions and is often considered a potential electrophysiological correlate of SI and LE.

However, due to their subtle and overlapping nature, tau rhythms remain challenging to isolate and are more elusive in EEG data than other sub-alpha rhythms such as mu oscillations, making them one of the less-explored subtypes of alpha oscillations in electrophysiological research.

Unlike occipital or sensorimotor alpha rhythms, tau oscillations are more elusive in scalp EEG recordings (Wisniewski and Zakrzewski, 2023). This is mainly because stronger alpha sources from occipital and parietal cortices tend to mask them (Weisz et al., 2011). As highlighted by Wisniewski et al. (2024), tau rhythms are more easily observed with MEG or fMRI, while EEG detection often requires advanced signal decomposition methods.

beta β (13-30 Hz) Also described by Berger (1929), beta rhythms are observed in motor, somatosensory, visual, and olfactory cortical areas (Beppi et al., 2021). They are believed to originate in the basal ganglia and project to the cortex *via* the thalamus, and also occur in deeper structures (Liu et al., 2020).

gamma γ (>30 Hz) Reported by Jasper and Andrew (1938), gamma oscillations occur in frontal, parietal and occipital lobes and are thought to play a role in integrative brain functions (C. S. Herrmann et al., 2004). They are challenging to analyze with EEG because they are easily contaminated by eye and muscle movements.

ERD, ERS and Inhibition Theory

Event-related desynchronization (ERD) and event-related synchronization (ERS) are fundamental concepts for interpreting oscillatory brain activity. They describe, respectively, a decrease or an increase in rhythmic synchronization of neuronal populations in response to a stimulus or task.
Alpha-band activity, in particular, has been closely linked to these mechanisms. As explained by Klimesch et al. (2007), "Alpha ERS plays an active role in inhibitory control and timing of cortical processing, whereas ERD reflects the gradual release of inhibition associated with the emergence of complex spreading activation processes."
In other words, alpha ERS is thought to reflect increased neural inhibition of the cortical regions, whereas alpha ERD reflects a decrease of neural inhibition, indicating active processing in the relevant cortical areas.
These dynamics are not limited to the alpha band but are observed across different frequency ranges, depending on the cognitive process and brain region involved.

Conceptual Clarification

In this project, we refer to cognitive inhibitory control, an executive function, and neural inhibition, reflected by alpha event-related synchronization (ERS). Although these concepts share similar terminology, they represent distinct mechanisms. Cognitive inhibitory control is detailed in Section 1.1 of Chapter 3 and the inhibition hypothesis in Box 2.2 of Chapter 2.

2.3 Interferences and preprocessing

Interferences

As with other physiological signals (ECG, EOG), EEG recordings result from the activity of an extensive number of sources. These sources are responsible for the amplitude, shape, and frequency observed during task or resting-state activities. Currently, it is not possible to precisely isolate and identify all the sources contributing to the global recorded signal. In addition to brain activity, some non-neural signals can interfere with EEG recordings. Common sources of interference include eye movements, cardiac activity, and muscle contractions. For instance, ocular motor activity alters the eye's electric field, which diffuses across the body.

Due to the proximity of the eyes to the prefrontal cortex, these fields will impact the signal, particularly in prefrontal electrodes. Blink-related peaks and saccade-related deflections (squared-shaped artifacts) appear in specific electrode locations (*see figure interference*.

Motor activities, especially maxillofacial movements, can introduce large and complex artifacts. These are often more difficult to manage than eye movement artifacts, as they tend to be irregular, non-repetitive, and non-periodic (*see figure interference*). Other forms of interference may arise from environmental sources, such as power line noise. In Europe, for example, line noise occurs at 50 Hz and can strongly affect the signal in that frequency band. However, this type of interference can be easily removed using a notch filter (see Box 2.16).

Preprocessing

Due to the presence of interferences and noises, raw EEG data must be preprocessed prior to analysis. Preprocessing is an essential step for cleaning the data and being able to extract information from it. However, excessive preprocessing can distort the signal and bias results (Delorme, 2023). A major challenge in EEG preprocessing is the need for visual inspection of the data; some portions of the data often need to be manually removed.

A typical preprocessing pipeline includes the following steps: re-referencing (based on a specific channel or the average signal), filtering to remove very low or high frequencies and exclude unwanted bands, resampling to reduce data size, and adjusting the sampling rate. Often, a channel localization is required to ensure that each channel is correctly mapped to its corresponding scalp electrode.

In order to remove residual or irrelevant signal components, a common approach is to use independent component analysis (ICA) during preprocessing to isolate and remove artifactual components. After basic preprocessing, it is normally possible to observe basic EEG results, such as for example the alpha oscillations when an individual closes his or her eyes.

Filters in EEG

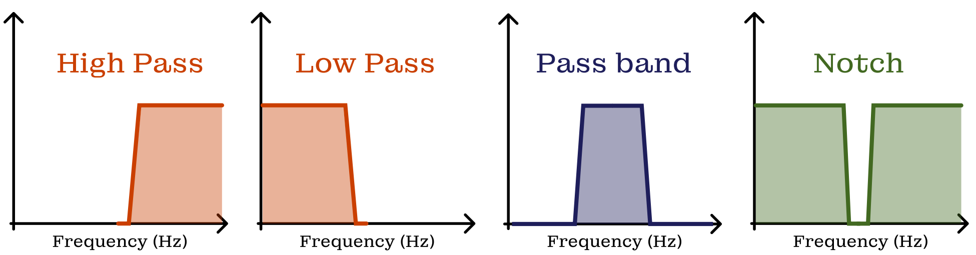


Figure 2.16: Filter types. Filtered data are represented in color for each filter type.

Independent Component Analysis

Consequently, the recorded EEG data is the result of electrical activity generated by unmeasurable sources within the brain and body, resulting in a mixture of components. The number of EEG electrodes (n), placed at different scalp locations, may vary depending on the desired spatial resolution. Each electrode records a projection of the same global brain activity from a different position on the scalp.

Independent component analysis (ICA) is a computational method used to separate this mixed signal into statistically independent components. Applying ICA to EEG data produces n components, each represented by a topographical activity map. These components can be classified based on their similarity to known EEG patterns.

The extracted components are often visualized as potential scalp maps and can be ordered by the global explained variance, for example. One of the main purposes of ICA is to distinguish components originating from brain activity from those generated by non-neural sources, such as eye movements, muscle activity, and electrical noise (e.g., line noise). The number of extracted components corresponds to the number of recording channels.

Once unwanted components are identified, they can be removed, and the EEG signal can be reconstructed using the remaining components. This results in a cleaned, albeit reduced, signal, with major interferences excluded. However, visual inspection is still recommended to ensure signal quality.

Cocktail party problem - Independent Component Analysis

A simple example of cocktail party effect. In a room, multiple people are having conversations, we will call them *speakers*. A microphone is installed in the center of the room and records conversations. The signal recorded by this microphone is a single sound waves indicating the results of all sound inputs originating from the different speakers, and the environmental sounds and noises. By listening at the microphone output only, the human brain is able to isolate one of the speakers in order to listen to his or her statements only.

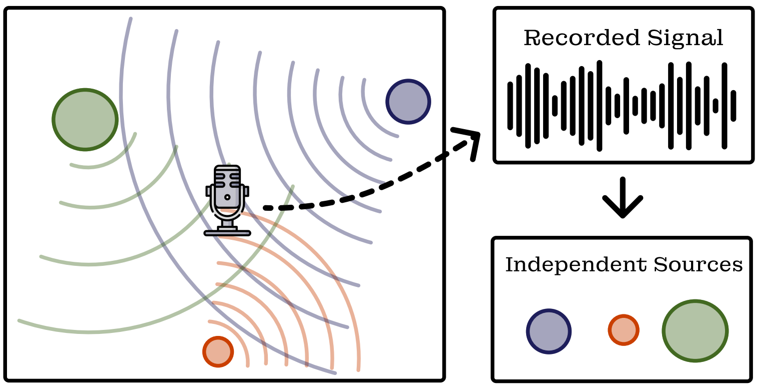


Figure 2.17: Cocktail party problem and independent sources extraction.

2.4 Analyses

Once the EEG signal is recorded and preprocessed, different types of analysis are possible, depending on the experimental design and hypothesis. In this section, we briefly describe the most common analysis methods as well as some more specialized one, such as microstates analysis or components analysis.

Time domain

The high temporal resolution of EEG allows analysis of neural activity at the millisecond scale. Event-related potentials (ERPs) reflect a stereotypical brain response to a given stimulus. This kind of analysis represents one of the most commonly used approaches in the time domain and is widely documented (Luck, 2014; Sur and Sinha, 2009). ERPs are locked to a stimulus onset (either exogenous or endogenous). Because one stereotypic response has a low SNR (1 to 10 or 1 to 100), repeated presentations of the same stimulus are needed to obtain an averaged neural response with a better SNR, all other activities unsynchronized with this onset will be statistically subtracted.

However, while ERP analysis is informative for short auditory or visual stimuli, its application to continuous speech is more challenging. Speech occurs on the scale of seconds, whereas ERPs capture activity on the millisecond scale. Although ERPs can still be applied in speech research, alternative approaches such as time-frequency or component analyses may be more adapted for the study of speech perception.

Frequency domain

Spectral analysis of the EEG signal is a simple and accessible analysis type that quantifies the power of specific frequency bands. It relies on the assumption that the signal is stationary within short time windows, meaning that the mean and variance are not time-dependent.

To convert EEG data from time to frequency domain, a Fast Fourier Transform (FFT, see Box 2.18), among other tools, can be applied. The signal is decomposed into its oscillatory components, represented by an infinite sum of sinusoids with different amplitudes and phases. The resulting power spectrum reflects the distribution of power across frequencies and can then be interpreted according to the experimental conditions.

Fourier Transform

Fourier transforms convert a signal from the time domain to the frequency domain. It breaks down a signal (e.g., EEG signal) into sinusoidal components defined by their frequency, amplitude, and phase. The sum of all extracted sinusoids permits the recovery of the original time domain signal, with an inverse Fourier transform.

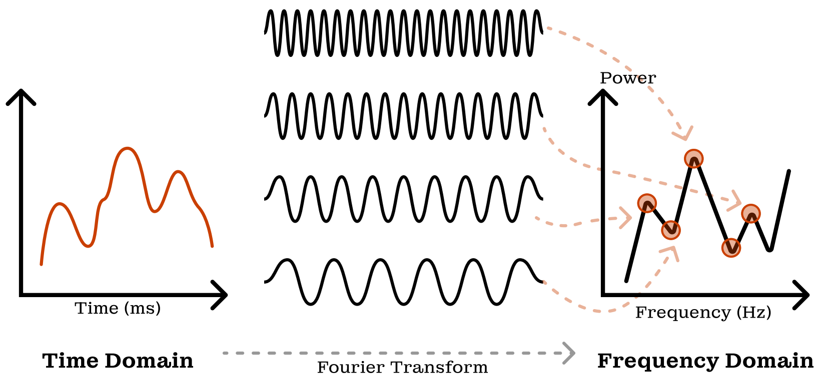


Figure 2.18: Time-to-frequency domain conversion using Fourier transform.

Time-frequency domain

Time or frequency domain analyses consider only one dimension at a time, either temporal or spectral. Time-frequency methods combine both, allowing the observation of oscillatory activity through time.

Event-related spectral perturbations (ERSPs) represent changes in spectral power over time relative to a reference period. To extract ERSPs, the EEG signal is decomposed into frequencies within successive time windows, using Morlet wavelet or FFT. As explained by Makeig (1993) ERSP "measures the average time course of relative changes in the spontaneous EEG amplitude spectrum induced by a set of similar experimental events."

Spectral changes over time can be quantified as either absolute power (event-related synchronization, ERS) or relative power (ERSP); the difference stands in baseline corrections. For ERSP, the average power before the stimulus is computed and subtracted or divided from all time windows. This allows a better visualization and highlights the stimulus-related modulations of the spectral activity.

However, interpreting ERSPs is not always straightforward. A single peak may reflect the combined influence of several neural processes and sources projecting to the recording electrodes, making the outcome complex and difficult to interpret. Makeig et al. (2004) describe this concern well when they report: "a single peak or valley in an ERSP image might index the combined effects of multiple processes modulating the activities of several EEG sources that project to the recording electrode."

Components analyses

In another approach to EEG analysis, aspects beyond the time and frequency domains can be examined. For example, ICA (see Section 2.3) can be used to determine characterized components of the signal.

Time and frequency analyses, such as ERP and ERSP, may not fully capture the cortical dynamics present in the EEG data. These measures overlook two important issues: the spatial mixing of multiple cortical sources and the role of phase resetting (Makeig et al., 2004). The recorded activity is the combination of signals from multiple sources in which positive and negative potentials occur. Linear decomposition methods, particularly ICA, address these limitations by separating EEG into independent components, allowing the application of source-specific analysis. This approach provides physiologically more plausible oscillatory dynamics, more direct links to behavior (Makeig et al., 2004), and enables trial-by-trial and group analyses.

During the study of electrophysiological indices related to speech and complex auditory processing, ICA can be used not only for preprocessing purposes but also for deeper analysis beyond the time-frequency domain.

In a group study, the individual ICA components can be compared across participants to identify common patterns of brain activity depending on the experimental conditions. To do so, components from all participants are clustered (using a k-means algorithm, for example) into a given number of clusters (k). Each cluster is represented by a mean map and includes a set of components across subjects. The cluster centroids can then be identified based on other component analysis.

For example, if for N subjects, which data were recorded with a 64-electrode-EEG, 64 x N components are computed. The 64 x N components can be clustered into k groups. Clusters that include components from a large number of participants are generally of greatest interest, as they may reflect more consistent, population-level brain activities. Certain clusters in particular may be associated with specific known brain activity, such as tau rhythms, for example. Furthermore, each component is associated with an estimated source of activity. The location can be visualized as the centroids of the clusters, offering further insight into the anatomical organization of brain activity.

Source localization - The inverse problem

Solving the inverse problem (see Box 2.19) in EEG is a major challenge, as many different source configurations can produce similar electrical fields. To reduce this ambiguity, assumptions are made about the spatial and temporal distribution of neural activity (Michel and Brunet, 2019). In many cases, an assumption about the number of sources is made to decomplexify the inverse problem, suggesting that only a small number of sources (often one) is the origin of the generated signal of interest. Although such constraints make the problem more tractable, they can bias source localization by overlooking sources or generating unrealistic sources. Another limitation is the anatomical variations of the head. The electrical properties of brain tissue and the skull thickness are not homogeneous (Michel and Brunet, 2019), and along with the head shape, they influence the signal propagation. Therefore, structural MRI can be used for better estimation of the head geometry and electrode positions, as the accuracy of the lead field determines the precision of source localization (Michel and He, 2019).

The inverse problem

In EEG signal processing, source localization relies on two complementary problems: the forward and the inverse problem. The forward problem assumes that the neural sources are known and aims to compute the resulting electrical potentials at the scalp sensors. This problem has a unique solution, as it describes how activity propagates from the brain to the electrodes (He and Lian, 2005).
In contrast, the inverse problem consists of estimating the neural sources that generated the recorded scalp potentials. For a given scalp distribution, there is an infinite number of possible source configurations (in terms of number, size, orientation, and power of dipoles). Because there are infinite configurations, the inverse problem is a major mathematical and physiological challenge, requiring anatomical and spatial constraints to achieve plausible solutions (He and Lian, 2005).

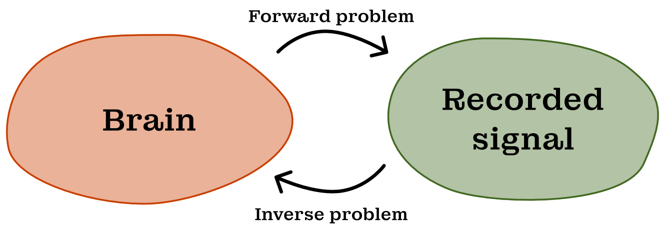


Figure 2.19: Inverse and forward problems.

Multiple algorithms and methods are available; one widely used method in EEG signal processing is the low resolution electromagnetic tomography (LORETA, Pascual-Marqui et al., 1994), which assumes that neighboring neurons have a simultaneous and synchronous activity (Pascual-Marqui et al., 1994). For more details on the different methods for source localization, please refer to Grech et al. (2008), He and Lian (2005), He et al. (2018), and Michel et al. (2004)

Microstates

The brain is continuously active, even in the absence of external stimulation. Resting-state activity is organized into spatiotemporal networks that generate intrinsic dynamics described by Deco et al. (2011) as a constant inner state of exploration. This ongoing activity provides insights into the global patterns of the brain and can be considered as an indicator of mental state, complementary to the other methods mentioned previously.

The brain at rest can be observed through different neuroimaging methods. fMRI has revealed intrinsic resting-state brain properties, but its low temporal resolution makes it less suitable for capturing the rapid changes of mental activity. With its high temporal resolution, EEG is more adapted for the study of fast-changing resting-state dynamics (Michel and Koenig, 2018).

The spontaneous topographic activity recorded by EEG is dynamic. EEG microstates, first described by Lehmann et al. (1987), are brief periods during which the scalp potential topography remains (pseudo) stable. At rest, these consecutive states last between 60 and 120 ms (Michel and Koenig, 2018). Microstates are transient activation maps that spontaneously switch from one configuration to another. These maps reflect the global coordination of neuronal activity over time, meaning that their change indicates a reorganization of brain networks (Michel and Koenig, 2018). Because they fluctuate between reproducible and organized patterns, microstates provide a framework for analyzing brain activity at rest, and possibly, although more challenging, during tasks.

Microstates Definition

Michel and Koenig (2018) define the microstates as "global patterns of scalp potential topographies recorded using multichannel EEG arrays that dynamically vary over time in an organized manner".

Those activation maps are consistent across healthy individuals. At rest, any EEG signal can be classified into common prototypical maps. Microstates activity is continuous: at every moment, the topographical activity of the scalp can be associated with one of these maps. This reproducibility allows microstate analysis at both the individual and group levels. Group-based studies, in particular, have provided insights into the general networks associated with specific tasks or cognitive functions.

The prototypical maps that can be extracted from group analyses can also be applied to individual data, making it possible to classify and compare microstate metrics across subjects or experimental conditions. The activation pattern of microstates is relatively stable across participants, ensuring good intra-subject reproducibility.

Microstates Metrics

Microstates can be quantified using several metrics. The most common ones include:

Coverage proportion of time a given microstate is present within a time window.

Duration length of time each microstate stays stable in a time window.

occurrence frequency of appearance of a microstate in a time window.

Canonical Microstates For many years, four microstates were commonly described as the canonical microstates. The canonical microstates (see Figure 2.20) were used as reference for classification and labeling of the prototypical maps in the microstate analysis pipeline (see Section 3.2 in Chapter 5). These were associated with different resting-state networks including auditory (microstate A), visual (microstate B), executive control (microstate C), and attentional (microstate D) systems (Michel and Koenig, 2018). However, more recent research has questioned this fixed classification, suggesting that restricting analysis to four canonical microstates may limit interpretation (Tarailis et al., 2024). Instead, methodological approaches such as the global explained variance (GEV) are recommended to estimate the optimal number of prototypical microstates for a given dataset.

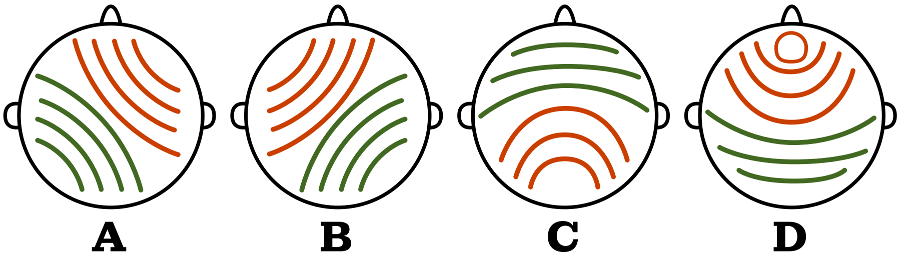


Figure 2.20: Canonical Microstates (k=4).

3 Chapter conclusion

Speech perception involves multiple levels of processing, from auditory encoding in the ear to integration in the auditory cortex and higher cortical areas. Listening effort, in particular, is reflected in cortical activity, notably in the IFG.

In complex auditory situations, speech perception is multimodal and relies on the interaction of several cognitive functions. While research on LE and SI often focuses on specific aspects of this process, adopting a broader, domain-general perspective provides a more comprehensive understanding. Executive functions, as high-level cognitive mechanisms, play a role in extracting relevant information from complex auditory scenes. These functions, and their relation to LE, will be defined and explained in the next chapter.

Summary of the chapter

- Sound is transferred to the brain by being transduced into neural signals in the inner ear.
- Primary auditory pathways carry these signals from the cochlear nuclei to the auditory cortex.
- Cortical pathways integrate auditory information and enable higher-level processing.
- The IFG, particularly the left IFG, plays a key role in both speech processing, especially under adverse conditions.
- EEG provides a non-invasive physiological measure that is used to investigate indices of LE.
- Alpha and theta oscillations in EEG may be associated with listening effort.
- Advanced analyses, including ICA and microstate analyses, allow exploration of neural correlates of listening effort beyond the standard time–frequency domain.

3

Executive Functions

Description of the Chapter

After exploring the brain pathways involved in speech understanding, we turn to the higher-order cognitive processes that guide how we perceive and respond to our environment. Executive functions enable listeners to focus attention, manage competing information, and adapt their strategies when faced with challenging listening situations. In this chapter, we describe the main models of executive functions proposed in the literature and detail the selection of the model used for this project. Additionally, the brain regions and networks that support them, as well as their relationship with listening effort, are presented in this chapter.

1 Executive functions

Executive functions (EFs) refer to a set of interrelated high-level cognitive processes that support goal-directed behavior and adaptive functioning in daily life. Although the concept remains difficult to define, EFs include processes such as planning, problem-solving, attentional control, working memory, and decision making (Cristofori et al., 2019; Diamond, 2013).

Historically, EFs have been associated with frontal lobe activity, particularly following the well-known case of Phineas Gage, who suffered from strong behavioral and personality changes after important frontal lobe damage in 1848 (see Box 1.2). Following neuropsychological research confirmed the involvement of the frontal cortex in executive functioning. More recent findings, however, suggest that EFs brain activity is not exclusively related to frontal activity but also involves subcortical networks. This complexity underscores the need for clarity and precision in the definitions and terminology used when discussing EF.

The term EFs has evolved over time. Pribram (1973) referred to the ”executive” role of the frontal cortex in the organization of behavior, while Lezak (1982) first talked about processes responsible for regulation, control, and execution of goal-directed actions using the term ”executive functions”. Lezak also emphasized the distinction between executive and cognitive functions, the latter referring to the extent of an individual’s knowledge or intellectual abilities. Chan et al. (2008) further described EF as an ”umbrella term” encompassing multiple processes, including inhibition, planning, working memory, cognitive flexibility, initiation, and monitoring. This last notion is often used in the literature to describe EFs. Various theoretical models have been proposed to describe EF from neuroanatomical, psychological, and statistical perspectives. Among the most widely accepted contemporary models (Diamond, 2013; Miyake et al., 2000), EFs are typically characterized by core components: updating of working memory, cognitive flexibility, and inhibition. The following section outlines these models in greater detail. In this thesis research, we were particularly interested in inhibition and its relation to listening effort.

1.1 Executive function models

Luria (1966)

A first neuroanatomical model of EF was proposed by (Luria (1966)), which described how some brain structures interact to support complex cognitive tasks. Through observations of patients with frontal lobe damage, Luria (1966) identified a relationship between frontal lobe activity and behaviors now associated with EF. These patients often exhibited dysexecutive symptoms, such as difficulties adapting their behavior based on environmental feedback, highlighting the role of the frontal lobes in goal-directed regulation and cognitive control.

Baddeley & Hitch (1974)

Revised in 2000 (A. Baddeley, 2000), the working memory model is composed of four subsystems: the phonological loop, the visuospatial sketchpad, the episodic buffer, and the central executive. This model proposes a separation between visual and auditory encoding in short-term (or working) memory, represented by the two subsystems: phonological loop and visuospatial sketchpad (see Figure 3.1). The central

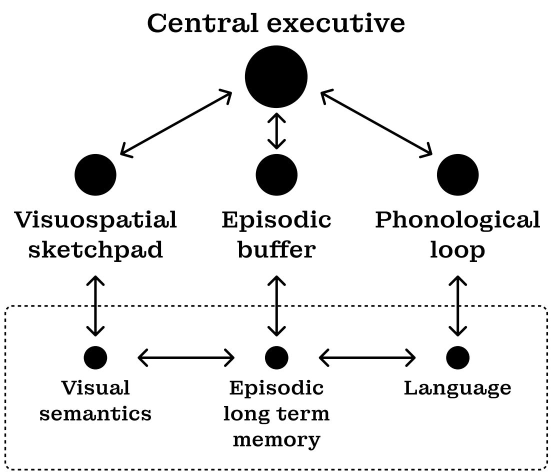


Figure 3.1: Executive function model - adapted from A. D. Baddeley and Hitch (1974).

executive supervises these subsystems, coordinating their activities and allocating attention. The episodic buffer acts as a temporary store that integrates information from the phonological loop, visuospatial sketchpad, and long-term memory, facilitating the creation of coherent episodes or experiences. Although the central executive is not anatomically localized in this model, it is conceptually linked to core executive functions such as inhibition, task switching, and updating, processes that are central to goal-directed behavior and cognitive flexibility.

Norman & Shallice (1986)

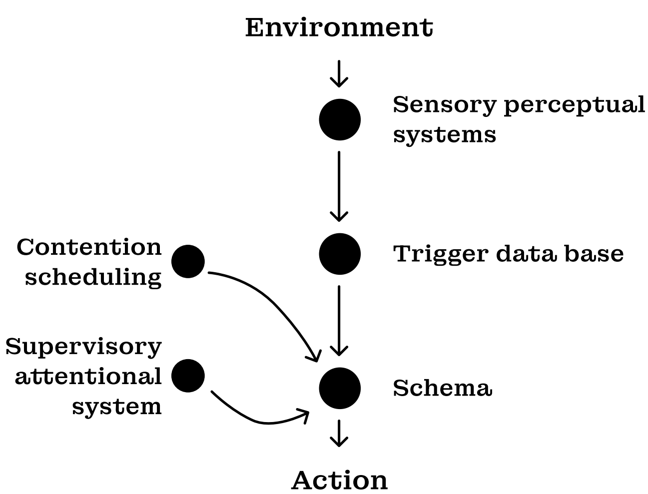


Figure 3.2: Executive function model - adapted from Norman and Shallice (1986).

Norman and Shallice (1986) model of executive functioning distinguishes between known and unknown situations to explain how cognitive control operates in the selection and regulation of behavior. In familiar context, the sensory-perceptual system processes incoming environmental information, such as visual or auditory stimuli, and rapidly activates known responses by selecting appropriate action schema. These schemas are stored in the trigger database, which contains structured knowledge about sequences of actions that have been previously learned and automatized. However, when facing unfamiliar or complex situations that cannot be resolved through automatic responses, the supervisory attentional system (SAS) is engaged. It is responsible for deliberate control processes, including the formulation of new strategies, the inhibition of inappropriate responses, and the planning of novel sequences of action. The contention scheduling mechanism is an intermediary system to avoid conflicts between competing automatic schemas. It regulates behavior by suppressing irrelevant or conflicting responses and can recruit the SAS if necessary. Motor actions are selected and initiated according to the selected schema and output of the contention scheduling system.

Stuss & Benson (1984)

The executive function model proposed by Stuss and Beson (1984) is based on clinical observations of patients with brain damage. It conceptualized executive functions as emerging from a hierarchically organized system, composed of three levels: a lower cognitive function system responsible for basic processing, an intermediate supervisory system involved in the coordination of tasks and responses, and a metacognitive system, which enables self-awareness, goal setting, and strategic regulation. This model also distinguishes between the ”cold” aspects of EF, logical and abstract processing, and the ”warm” aspects, which include affective and social components.

In later work, Stuss and Alexander(2011) emphasized the limitations of attributing EF strictly to the frontal lobes, noting that early research often relied on patients with imprecise or diffuse lesions. He argued that EFs can also arise from damage outside the frontal cortex, reflecting the distributed and multifaceted nature of executive processing.

Miyake (2000)

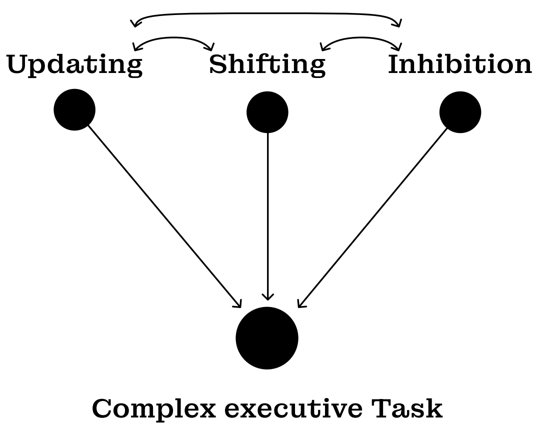


Figure 3.3: Executive function model - adapted from Miyake et al., 2000.

Miyake et al. (2000) aimed to investigate executive functions using a factorial analysis approach, in contrast to earlier models, often based on neuropsychological issues. Their goal was to observe the independence among the three mainly described EFs: shifting, updating, and inhibition. According to their model, these functions contribute differentially to higher-order executive tasks.

Shifting (also referred to as switching) refers to the ability to disengage attention from one task or mental set and transition to another. This process incurs a measurable cognitive cost, often reflected in slower reaction times or increased errors during task switching.

Updating involves monitoring and encoding of information in short-term memory, closely related to working memory.

Inhibition is the capacity to suppress automatic or dominant responses to a stimulus, allowing for goal-directed behavior. This relates to the ability to control a response directly related to a stimulus and overcome the automatization. A classical example, central to this project, is the cocktail party effect, in which one must inhibit irrelevant auditory input to attend to a specific voice in a noisy environment (Cherry, 1953).

To validate this theoretical model, Miyake et al., 2000 conducted a series of tasks designed to assess nine behavioral indicators related to the three core EF. Using latent variable analysis, they modeled inhibition, shifting, and updating as distinct but correlated constructs. Using these data, they extracted correlation coefficients to determine the degree of distinction among the three processes, and demonstrated that each EF contributes specifically to complex task performance while also sharing common cognitive resources.

Diamond (2013)

Diamond (2013) model builds upon earlier theoretical frameworks by refining core concepts such as working memory and inhibitory control. In addition, it introduces new components grouped under the umbrella of higher-level executive functions, including reasoning, problem-solving, and planning, related to fluid intelligence. A key feature of this model is cognitive flexibility, defined as the capacity to change perspectives or adopt new approaches when faced with changing task demands. This ability is particularly relevant when a problem requires abandoning an initial strategy in favor of a novel one. Such flexibility depends on the effective coordination of inhibitory control to suppress the previously dominant perspective and working memory, to maintain and manipulate alternative solutions. Diamond emphasizes that although these components are functionally distinct, they operate in a coordinated and hierarchical manner to support complex cognition and adaptive behavior.

1.2 Assessment of executive functions

Once executive functions are defined, the next step is to measure variables related to each sub-process to address specific scientific questions concerning EF. Regardless of the theoretical model considered, one major challenge lies in accurately measuring the variables related to each sub-process to address specific scientific questions concerning EF. These questions might directly involve EF, the processes they undergo, their variation throughout the lifespan, and environmental changes, other

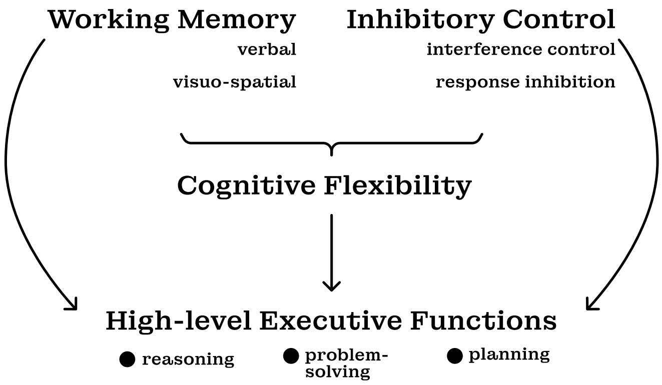


Figure 3.4: Executive function model - adapted from Diamond, 2013.

related cognitive functions, etc. The core problem remains the following: How do we measure EFs? Do we measure executive functions as a global process, using a single overall score derived from multiple sub-tests, similar to how IQ is assessed with the Wechsler Adult Intelligence Scale? Or do we independently measure each sub-process?

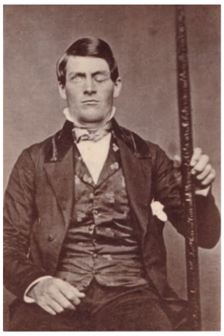
The complexity of measuring EF can be slightly reduced by selecting a single EF model and relying on the authors' descriptions of the tasks used for assessment. For example, following Miyake and Friedman's approach, we focused on three core processes (updating, switching, and inhibition), which, according to the authors, can be assessed by nine specific, validated, and documented tasks.

Miyake et al. (2000), through their model, choose to focus on three EFs that are regularly described in the literature, each assessed by three specific tasks with corresponding dependent variables. For updating tasks, the number of correct responses or recalled items was measured. For switching, the shift cost, calculated as the difference in reaction time between switch and non-switch trials, was used. For inhibition, performance was typically represented by the proportion of correct responses or the inhibitory cost, defined as the difference in reaction time between conditions.

However, laboratory measures may lack ecological validity. Performing tasks in a controlled, stimulus-free environment does not fully represent the complexities of everyday life. Yet, this control is intentional: the goal is to minimize unexpected variables to better understand what we are measuring. Although absolute certainty about the exact processes we observe is impossible, scientific rigor and determination allow us to gain meaningful insights into executive functioning.

Phineas Gage

In 1848, Phineas Gage, a young railroad worker, sustained extensive frontal lobe damage when an iron bar was accidentally driven through his skull (Kandell et al., 2000). This case, later studied in detail by Harlow (1869) and subsequently re-examined by Damasio et al. (1994), became one of the earliest and most cited examples linking frontal lobe integrity to personality and executive functioning. Following the injury, Gage reportedly underwent profound behavioral and psychological changes, becoming emotionally unstable, impulsive, and socially inappropriate. His case laid the foundation for the understanding of the frontal lobe's role in executive control and social behavior.



1.3 Choice of the model

After reviewing the range of existing models, it becomes evident that EFs are difficult to define in a unified manner. Over time, research has led to the diversification of EF models, each emphasizing different processes, levels of analysis (neuroanatomical, cognitive, statistical), or practical applications. To provide clarity and consistency for this project, we chose to adopt the framework proposed by Miyake et al., 2000. This model identifies three core components: updating, shifting, and inhibition, which are among the most empirically supported and frequently cited in the literature (Baggetta and Alexander, 2016). It was acknowledged that Miyake's model may not fully account for more complex executive functions such as reasoning or coordinating multiple control processes, suggesting that the model is not exhaustive but rather a foundational framework. As noted by Baggetta and Alexander 2016 in their systematic review, there is no single, universally accepted definition of executive functions in the literature, and authors often rely on multiple frameworks to clarify their approach to EFs. To avoid confusion arising from differing conceptualizations, we base our approach on Miyake and Friedman's model, adopting a domain-general perspective in line with our aim to examine the role of EFs, particularly inhibition, in listening effort.

2 Brain regions and networks involved in EF

EFs are a neuropsychological concept describing human behaviors and actions. These constructs have effective related brain activity that researchers have been trying to describe since they started to look at the brain as a coordinator for human behavior and decision-making. In this section, we describe the brain structures that were described as related to executive functions, and then the relative brain network activity.

2.1 Frontal lobe

The frontal lobe, and more specifically the prefrontal cortex, has historically been central to the understanding of EF. Clinical observations of patients with frontal cortex damage have played a crucial role in identifying executive dysfunction, often referred to as ”dysexecutive syndrome” (Stuss and Beson, 1984). These individuals typically retain intact perceptual and motor abilities but exhibit specific impairments in planning, regulation of behavior, and emotional stability. Consequently, the prefrontal cortex is now recognized as a key structure in cognitive control, decision-making, and the flexible coordination of behavior.

Unlike bottom-up processes, driven automatically by the nature of sensory stimuli (see Box 1.5), the prefrontal cortex supports top-down processing (see Box 1.5), allowing cognition to guide actions in novel or complex situations (Miller and Cohen, 2001).

Although the diversity of executive processes suggests no single brain area can account for all aspects of EF (Kandell et al., 2000), the prefrontal cortex is consistently implicated across sub-processes described by Miyake et al. (2000) such as updating, closely related to working memory (Rottschy et al., 2012), inhibition (Zhang et al., 2017), and switching (Wager et al., 2004). More broadly, neuroimaging meta-analyses confirm that the prefrontal cortex is involved in a wide range of executive control processes (Zink et al., 2021). Nevertheless, linking anatomical regions to complex neuropsychological constructs like EF remains challenging due to issues of circularity in interpretation (Duffy and Campbell III, 2001; Duke and Kaszniak, 2000).

2.2 Network-based understanding of EFs

Early studies on brain activity related to EFs mainly focused on single structures, especially the prefrontal cortex. However, the brain is an organ that works as an ensemble and whose structures cannot operate independently. Therefore, studying EFs by isolating individual brain areas risks shortcuts and oversimplifications of the complex neural activity underlying these functions.

Regarding EFs specifically, the diverse processes involved in cognitive control tasks suggest that multiple brain regions contribute to it, forming a complex network (Kandell et al., 2000). While the prefrontal cortex plays a central role, it does not act alone. Numerous studies have identified networks and brain regions associated with EFs, notably the right ventrolateral prefrontal cortex, the anterior cingulate cortex, and basal ganglia (Zink et al., 2021). These structures interact within spatially distributed networks that include additional brain areas (Zink et al., 2021).

Is it common in neuroimaging studies to isolate a cognitive process, trying to associate it with a single brain area. Nevertheless, the modern tendency highlights that networks involved in EFs are more complex, encompassing more structures than early clinical literature suggested. Indeed, the size of a lesion does not reliably predict cognitive deficits, and EF dysfunctions do not automatically arise from a brain lesion. Zink et al. (2021) emphasizes that EFs are not controlled only by top-down signals or specific brain areas. Instead, they emerge from communication across many brain regions working together as a network to manage different aspects of EFs. Using the example of bird flock dynamics, they illustrate how a distributed control system, operating through simple rules and fast communication between items of the system, can function effectively without a central top-down controller.

Overall, the evidence supports a shift from strict localization toward understanding EFs as arising from connectivity and integration within broad neural networks.

To explore this view in detail, several meta-analyses have been conducted on the executive functions described by Miyake et al. (2000) (updating, shifting, inhibition), each highlighting distinct but overlapping neural networks associated with these processes.

Updating

Updating, often studied through the lens of working memory, refers to the storing of limited but accessible information for a short period of time. It involves keeping track of new information that is important for the current task and regularly updating outdated or irrelevant information with more useful content. Largely shaped by A. D. Baddeley and Hitch (1974) theoretical framework, working memory is understood as a short-term system responsible for holding limited but accessible information for brief periods. According to the meta-analysis of Rottschy et al. (2012), working memory consistently involves a bilateral fronto-parietal network across different task types and stimuli, revealing a robust and generalizable core network. Notably, no single brain area is solely responsible for the ”central executive” function of working memory. Within the dorsolateral prefrontal cortex, a functional distinction is observed; the anterior portion is more involved in task-set control, while the posterior portion is sensitive to memory load. Task-set effects are mainly seen in the left hemisphere, involving areas such as the frontal part of the lateral prefrontal cortex, the superior parietal lobule, the intraparietal sulcus, and the anterior insula. In contrast, load effects are linked to activity in both sides of the inferior frontal regions. Furthermore, specific activation patterns vary with the nature of the working memory task: verbal working memory tasks more consistently activate the left Broca’s area, while non-verbal tasks tend to recruit dorsal and medial premotor regions. Memory for stimulus identity engages the posterior inferior frontal gyrus, whereas memory for spatial location involves the posterior superior frontal gyrus.

Shifting

Shifting, also referred to as cognitive flexibility, is the ability to switch between tasks, mental sets, or strategies, and is one of the core EF identified by Miyake et al. (2000). It is closely interrelated with other executive processes, particularly inhibition and updating. Like inhibition, shifting is not a unitary process but occurs in various forms depending on the cognitive demands. As reviewed by Wager et al. (2004) in a meta-analysis of attention-shifting studies, several cortical regions were reliably activated across diverse types of shifting, including posterior areas and frontal regions. Common activation patterns emerged in the parietal, premotor, and medial prefrontal areas, regardless of the specific type of shifting. However, subregions within the right premotor cortex and the anterior and posterior intraparietal sulcus were found to respond differentially depending on the specific type of shifting. This suggests that while a core shifting network exists, distinct types of shifting may rely on partially separable neural mechanisms.

Inhibition

Inhibition, as defined by Miyake et al. (2000), refers to the ability to deliberately suppress dominant, automatic, or prepotent responses when necessary. It is a multifaceted EF that can be subdivided into distinct strategies and behavioral components, such as interference resolution, action withholding, and action cancellation. As highlighted by Zhang et al. (2017) in their meta-analysis, inhibition consistently engages a right-lateralized network of brain regions, including the IFG, the insula, as well as the median and paracingulate gyri. Moreover, the frontoparietal and ventral attention networks emerge as core systems involved across various inhibitory processes. Specifically, interference resolution (that can be illustrated by Stroop tasks (Stroop, 1935 for example) is associated with strong activation within the ventral attention network. Response inhibition is not a unidimensional construct, but comprises multiple subcomponents that engage both shared and distinct neural systems, depending on the specific cognitive demands involved.

3 Relationship between EF, LE and SI

EFs are often described as a crucial component of speech perception (Rudner and Signoret, 2016). Speech intelligibility and listening effort are multidimensional concepts that, in addition to sound perception processing, require higher-order cognitive functions. In complex auditory situations, different processes are involved, including the core EFs described by the model of Miyake et al. (2000). Among these mechanisms, updating of working memory plays a central role, as listeners must temporarily maintain speech information in order to integrate it into meaningful sentences. Shifting enables listeners to redirect their attention among competing conversations and segregate talkers. Finally, inhibition supports the listener to ignore irrelevant or distracting auditory input, such as masking talkers or noise.

Speech perception in complex auditory situations therefore, engages both auditory and non-auditory mechanisms (Eckert et al., 2016). Consistent with this view, stronger global cognitive abilities have been associated with better speech intelligibility and reduced effort (Strand et al., 2018). The specific links between each individual EF and LE could provide a better understanding of LE processes. Identifying precise links between EF and LE may ultimately provide valuable insights for predicting and assessing LE (Francis and Love, 2020).

3.1 Updating

Conceptual Clarification

Updating is a core executive function closely related to working memory, described by Miyake et al. (2000) as the ”monitoring of working memory representation”. It is therefore often studied within the scope of working memory. Although these two constructs are not identical, they are strongly intertwined. This distinction is not always made explicit in the literature, and it is important to note that findings attributed to working memory sometimes rely on experimental protocols that use tasks also employed to assess updating capacity. Also, updating has been characterized as one primary mechanism through which working memory contributes to psychological functioning (Carretti et al., 2005).

Working memory is the EF most frequently cited in research linking higher-level cognitive processes with listening effort and speech intelligibility in complex auditory environments. It has been shown to contribute to speech intelligibility (Besser et al., 2013; Francis and Love, 2020; Ingvalson et al., 2015 and may also reflect the associated listening effort (Rudner and Signoret, 2016), with a decrease of subjective LE with greater working memory (Stenbäck et al., 2021).

The Ease of Language Understanding model (ELU, Rönnberg, 2003; Rönnberg et al., 2013) has been proposed to explain the relationship between working memory and speech listening. It points out the importance of predictive processes during speech comprehension, which rely on working memory to integrate and anticipate linguistic input (Fedorenko, 2014; Lunner, 2003).

Physiological evidence also links working memory to speech comprehension (Francis and Love, 2020). For example, Wisniewski et al. (2017) frontal midline theta activity, especially power enhancement (Wisniewski et al., 2018), known to be associated with working memory (Onton et al., 2005), may also reflect how working memory is involved during effortful listening.

In addition to behavioral and neuroanatomical insights, it has been shown that training of working memory can significantly impact SIN intelligibility (Ingvalson et al., 2015).

3.2 Shifting

Shifting has been less extensively studied in relation to listening effort and speech understanding in complex auditory environments. Nonetheless, switching abilities play a role in speech comprehension (Perrone-Bertolotti et al., 2017), particularly in multi-talker situations, where the listener must repeatedly redirect their attention between talkers and reorient to the target talker (Lin and Carlile, 2015). Such shifts can occur either voluntarily or involuntarily (Koch et al., 2011), and contribute to the perceived LE (Brännström et al., 2018) by increasing the cognitive load (Best et al., 2008; Lin and Carlile, 2015).

3.3 Inhibition

In SIS or SIN situations, listeners must focus on a target talker to extract meaningful information. This requires the inhibition of masking talkers or background noise. Inhibition is at the core of this project, and its relationship with listening effort is not often discussed in the literature in comparison to working memory. However, it was shown that cortical activation patterns during inhibitory activity involves areas that also appear in LE-related activity. The IFG, especially the left one, shows modulated activity during both inhibitory control (Lanzilotti et al., 2022) and listening in complex auditory situations, especially in adverse conditions (see Section 1.2). Moreover, alpha power, often associated with LE, has been proposed as a marker related to inhibition (Klimesch et al., 2007; Wisniewski et al., 2017). Although overlap in activation does not imply identical processing, it may partially explain the observed links between inhibition and LE and could be considered as a potential explanation of their relationship.

In multi-talker situations, suppressing irrelevant speech streams allows the listener to better understand the target talker. Lower inhibition abilities may increase LE by making the listener try to process both target and masker streams simultaneously (Perrone-Bertolotti et al., 2017), whereas stronger inhibition capacities are generally associated with better speech intelligibility (Stenbäck et al., 2016). This suggests that better inhibition supports masking release and reduces LE. However, Brännström et al. (2018) reported contradictory results, showing that listeners with greater inhibitory control reported higher LE under both quiet and noise conditions. Besides the fact that this information is counterintuitive, such findings reflect the interest and need for further investigation of the inhibitory control and LE relationship. Additionally, Lanzilotti et al. (2022) suggest that some listeners rely on sound level differences as a segregation strategy and that this strategy is based on inhibitory control. This could explain individual differences in performance. Disengagement could also contribute to such variability (see Section 2.2 and 2.4).

3.4 General domain

The brain operates as a highly interconnected system in which complex pathways and large-scale networks interact to support cognition. When studying specific cognitive abilities, such as inhibition or listening in challenging environments, specific brain regions and their interactions with broader networks are often analyzed. In language processing, particularly under adverse conditions, the recruited networks are not exclusive to language but overlap with some involved in other cognitive abilities (e.g., the involvement of the IFG in both inhibition and speech processing). This suggests that listening effort and speech intelligibility rely partly on domain-general mechanisms (MacGregor et al., 2022; Rodd et al., 2010). MacGregor et al. (2022) points out that there is a direct link between general domain and language comprehension in complex auditory situations. Thus, while bottom-up processes may be sufficient for speech comprehension, top-down and domain-general mechanisms may enhance their effectiveness (Fedorenko, 2014).

The executive control network, also called the multiple demand (MD) network (Duncan and Owen, 2000; Duncan et al., 2020; Fedorenko, 2014; MacGregor et al., 2022, see Figure 3.5) encompasses fronto-parietal regions engaged in executive control and EFs. Its contribution and necessity to language comprehension are still unclear (MacGregor et al., 2022). According to Fedorenko and Shain (2021), language-specific networks may work together with the general-domain MD networks, but may also be self-sufficient. In addition, MacGregor et al. (2022) explains different points of view regarding the relation between the MD network and speech comprehension and suggests that the whole MD network is not necessary, as some

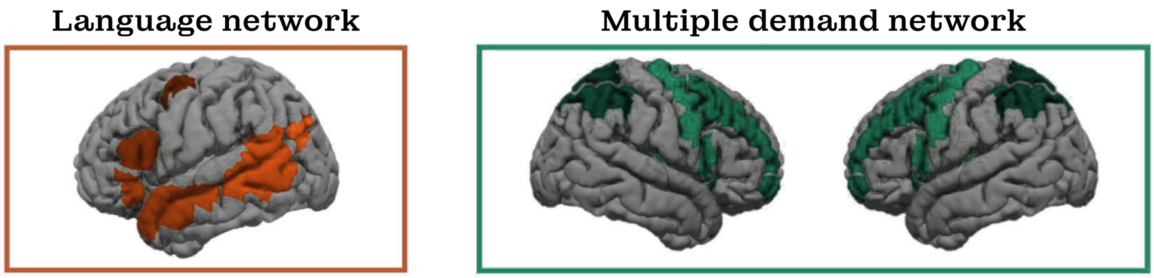


Figure 3.5: Language and multiple demand networks - adapted from Fedorenko and Shain (2021)

language comprehension is possible without some components of the MD.

Domain-general processes appear to be engaged when speech comprehension becomes effortful and cognitive demand increases. Thus, activity in the IFG, especially the left IFG, that plays a role in both language processing and inhibitory control, illustrates the overlap of neural resources across mechanisms (Rodd et al., 2010). This overlap shows how brain networks may be involved in different aspects of cognition, possibly interacting even if they are functionally distinct. Findings showing left IFG activation during complex sentence processing (Ye and Zhou, 2009) further support its role as a key region in the domain-general aspects of speech comprehension (Novick et al., 2010).

Overall, listening effort and speech intelligibility rely on both specific and general-domain networks. This overlap shows the relevance of investigating how LE and SI are related to EF and how they can be explored.

4 Chapter Conclusion

EFs are high-level cognitive processes that enable goal-directed and adaptive behavior. These processes have been described in the literature using various models, notably the one by Miyake et al. (2000), which identifies three core EFs: updating, shifting, and inhibition.

These high-level processes play a role in general cognition, but also in speech perception. In particular, the relation between listening in complex auditory situations and working memory is often explored in the literature. Although the relationship with inhibition is less frequently discussed, it shares common neural and behavioral mechanisms with listening effort (Lanzilotti et al., 2022; Stenbäck et al., 2016).

Understanding these relationships sets the stage for the next chapter, which explores cognitive training interventions targeting EFs such as inhibition, to potential enhancements of related cognitive functions.

Summary of the chapter

- Executive functions' model of Miyake describes three main core functions: updating, shifting, and inhibition.
- The prefrontal cortex is empirically related to cognitive control and executive functions.
- Different brain networks are associated with EF, notably ventrolateral prefrontal cortex, anterior cingulate cortex and basal ganglia.
- Inhibition and listening effort share overlapping behavioral and neural mechanisms.

4

Cognitive Training

Description of the Chapter

Building on our understanding of executive functions and their role in listening effort, this chapter explores cognitive training. We examine various approaches to training of executive functions, with a particular focus on inhibition, which is thought to play a key role in filtering irrelevant information in complex auditory environments. The chapter reviews different aspects of cognitive training, such as the control group, the placebo effect, and transfer effects.

1 Cognitive Training

Cognitive training refers to experimental designs that aim to enhance general or specific cognitive abilities through training (Gobet and Sala, 2023). Protocols may assess improvements in trained or untrained abilities, the latter assumed to be related to the trained ones via transfer learning (see Section 1.1).

The theoretical basis of cognitive training lies in neural and cognitive plasticity, which allows the adaptation of behavior, especially in complex environments (Strobach and Karbach, 2021; Taatgen, 2021). Training interventions exploit this plasticity by repeatedly engaging cognitive processes, with studies showing dynamic changes in brain activity (e.g., Eisner et al., 2010; Lanzilotti, 2021; Qi et al., 2019). Additionally, the Cognitive routine framework (Gathercole et al., 2016; Karbach and Kray, 2021) suggests that training leads to the development of new, automated cognitive paths when existing mechanisms are insufficient. However, despite these promising mechanisms, Taatgen (2021) explains that the analogy of cognitive training with physical training is inadequate: the brain is not a muscle, and there is evidence for both successful and unsuccessful cognitive training.

Cognitive training is appealing for the enhancement of performance and clinical applications, but its benefits remain debated and criticized (Gobet and Sala, 2023). These benefits seem to depend strongly on the type of training, the cognitive functions targeted, and the experimental protocols used (Jaeggi et al., 2017).

As explained in the Chapter 2, cognitive function studies are often related to neuroanatomical measures. Thus, one way to assess cognitive training effects is to use neuroimaging techniques to measure how brain activity evolves. Regarding the IFG in particular, and speech intelligibility training, some studies showed decreased activity after training (Lanzilotti et al., 2022; Qi et al., 2019) while others showed increased activity (Eisner et al., 2010). These apparently contradictory findings can be interpreted within the framework of neural efficiency (Neubauer and Fink, 2009) and learning dynamics (Dehais et al., 2020). During the initial stages of training, greater cortical recruitment reflects the effortful engagement of cognitive resources. As performance improves and processing becomes more automatic, neural activity tends to decrease, indicating more efficient brain functioning.

Cognitive training has grown in popularity over the last decade, driven by technical advances and more sophisticated analyses of longitudinal data (Strobach and Karbach, 2021). Interest has also been fueled by the fact that cognitive and neuronal plasticity occur throughout the lifespan and by the concept of transfer learning (Strobach and Karbach, 2021), which is discussed in the next section.

1.1 Transfer learning

Many cognitive training studies rely on the idea that practicing a specific task can enhance untrained cognitive abilities, a concept called transfer learning.

Transfer Learning

Strobach and Karbach (2021) defines transfer learning as ”the improvement of specific cognitive abilities by the transfer of training of another cognitive ability.”

Transfer can occur in two main forms: near transfer and far transfer (Taatgen, 2021). The distinction lies in the proximity between the trained function and the acquired skill. Near transfer refers to improvements across closely related domains, for example, studying arithmetic to improve in geometry. In contrast, far transfer involves training where the transferred skills are slightly or even not related, for example, training arithmetic to improve in Spanish (Gobet and Sala, 2023).

A main difference between these two types of transfer lies in the evidence supporting their effects. Several studies and meta-analyses report no evidence of beneficial far transfer from cognitive training (Melby-Lervåg et al., 2016; Sala and Gobet, 2017; Sala et al., 2019, while others show possible near transfer effects (Melby-Lervåg et al., 2016; Sala and Gobet, 2017). Following this, Harris et al. (2023) suggested exploring mid-transfer effects and their potential benefits for the optimization of human performance.

Overall, the effects of cognitive training are widely debated, and some studies reveal general transfer effects, particularly when training targets processing functions, such as executive functions (Strobach and Karbach, 2021).

1.2 Executive functions training

EFs are often the target of cognitive training. They may be the direct focus of the training task or the expected domain to benefit from transfer. As described in Chapter 3, EFs consist of distinct but interrelated sub-functions, and improvement in one does not necessarily imply improvement in another (Karbach and Kray, 2021). However, transfer is more likely to happen when the trained and untrained functions are in closer order of processing.

Inhibition Training

Training targeting inhibition specifically is rare and less common than training focused on the global EFs (Chavan et al., 2015). Moreover, inhibition is often indirectly influenced by working memory training, as its underlying cognitive processes are intertwined (Karbach and Kray, 2021). Evidence for transfer effects of inhibition is, as in general for cognitive training effects, heterogeneous. Some studies report transfer effects from trained inhibitory tasks on food control (Houben, 2011), alcohol consumption (Houben et al., 2011, 2012), or gambling behavior (Verbruggen et al., 2012). In contrast, other studies did not find evidence for transfer learning effects of inhibitory training (Enge et al., 2014; Thorell et al., 2009). In addition, Zhao et al. (2018) showed that inhibition training transferred to children's executive functions but not to those of young adults.

There is a lack of information regarding the behavioral and neuroanatomical effects of inhibitory control training (Chavan et al., 2015), especially compared to other executive functions such as WM. However, since impaired inhibitory control is implicated in several clinical disorders, such as addiction (Bechara, 2005), compulsive disorder (Bannon et al., 2002), or attention deficit hyperactivity disorder (Barkley, 1997), the potential clinical interest of training this specific executive function is considerable.

1.3 The control group

Using a control group and defining it precisely is essential for designing an experimental protocol. The control group consists of randomly selected participants who, depending on the study's ethical criteria, may not be informed of their group assignment. In double-blind studies, even the investigator remains unaware of each participant's assigned group. A well-designed experimental protocol should include such a group with as many participants as in the other groups. Participants in the control group may still exhibit changes due to expectations or engagement, known as the placebo effect (Foroughi et al., 2016) or the Hawthorne effect (Adair, n.d.).

Placebo Effect

The placebo effect is well documented in clinical and drug research, although its underlying mechanisms remain unclear. In the domain of cognitive training, its role has not yet been evaluated, raising additional questions about the validity of reported benefits (Foroughi et al., 2016). As a result, caution is advised when interpreting findings, and claims from the brain training industry should be considered with care (Foroughi et al., 2016).
In addition, the Hawthorne effect refers to a change in participants' behavior that occurs when they become aware that they are being observed(Adair, n.d.).

The control group can take different forms depending on the experimental design. A main characteristic is whether the control group is active or passive during the training period. A passive control group does not undergo any type of training and has as little interaction with the researcher as possible during the training phase. These participants only complete pre- and post-training sessions. Although this approach minimizes external influence, it does not control for placebo effects, motivation, or engagement differences between groups.

In contrast, an active control performs an alternative task, not related to the cognitive abilities of interest (Au et al., 2020). This design helps to account for non-specific factors such as task engagement or researcher interaction, but can be challenging to implement, as the control activity must be sufficiently distinct from the training task while still comparable in terms of cognitive demand (Gobet and Sala, 2023; Simons et al., 2016). This means that using a totally unrelated task, for example, replacing an arithmetic training task with a completely unrelated creative activity, is considered by Simons et al. (2016) as an insufficient active control condition.

Good practice for cognitive training

Several criteria have been suggested to reduce biases and increase chances of training success (Karbach and Kray, 2021; Zhao et al., 2018):

- Focus on higher-order cognitive functions rather than task-specific.
- Aim for near transfer effects by using training and transfer tasks that are closely related.
- Ensure tasks are sufficiently challenging, to avoid boredom or disengagement, which would lead to limited improvement.
- Employ adaptive training that gradually increases in difficulty or varies along the sessions.
- Conduct training across multiple sessions to allow consolidation and lasting effects.
- Use an active control group if possible.

2 Chapter Conclusion

Cognitive training explores how practicing specific tasks can modulate cognitive abilities and potentially produce broader benefits. It has faced considerable criticism, making a well-structured experimental design essential to ensure valid and reliable results.

In this project, our objective is to examine how inhibition training might affect listening effort and speech intelligibility performances. The question of near and far transfer is particularly relevant for this project, especially regarding how closely inhibitory control is related to LE. Listening in complex auditory situations involves diverse cognitive processes, with inhibition suggested as a key component (Lanzilotti et al., 2022; Stenbäck et al., 2016). In particular, in multi-talkers situations, where, after segregating the auditory streams, the listener must ignore the masker's stream to focus on the target talker, probably using inhibitory control mechanisms.

Summary of the chapter

- Cognitive training is an experimental design aiming to improve targeted cognitive abilities
- Transfer learning refers to the improvements in untrained tasks
- Executive functions are often the focus of cognitive training
- Inhibition is rarely the center of cognitive training and deserves further exploration
- The control group and the choice of the control task are crucial for ensuring the validity of the findings

III

Scientific Contributions

5

General Methods

Description of the Chapter

In this chapter, we describe the general methodology used across the three scientific contributions of this project. Specific methods related to each study or contribution are detailed in their respective sections.
In parallel, methodological adaptations and analysis pipelines developed as part of the project are presented in this chapter. These adaptations constitute methodological contributions in themselves, providing tailored tools and processing strategies to address the specific requirements of the projects. In particular, the EEG analysis pipelines are flexible and can be applied to other experimental designs beyond the present project.

1 Behavior

1.1 Speech Intelligibility - Coordinate Response Measure Corpus

The Coordinate Response Measure (CRM) corpus was used in this project to create different listening conditions. The CRM is composed of recorded sentences deriving from a model with 3 parameters: a call sign, a color, and a digit. The prototypical sentence is always the following: "Ready [Call Sign], go to [Color] [Digit], now". The Call Sign can be extracted from one out of eight possible (e.g., Arrow, Barron, Charlie, Eagle, Hopper, Laker, Rigo, Tiger), the color can be blue, red, white, or green and the digit a number selected from 1 to 8 (see Table 5.1 and Figure 5.1).

The French language version of the CRM has been designed to mirror, as closely as possible, the English language version (Isnard et al., 2024). The sentence structure remains almost identical: "*Prêt* [call sign], *va au point* [color] [number] *go*" with equivalent lexical variables (see Table 5.1). In the French version, the color *white* (blanc) was replaced with *yellow* (jaune) due to the phonetic similarity between the words "bleu" and "blanc" in French.

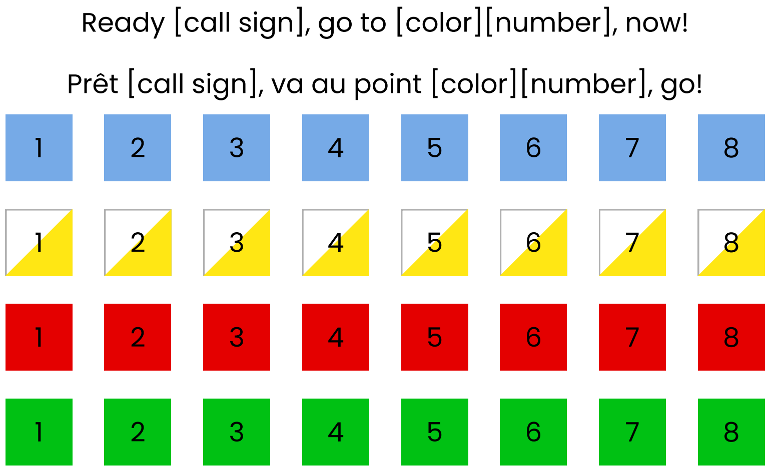


Figure 5.1: CRM typical sentences in French and in English with the interfaces. In English, the color of the second row is yellow, and it is white in French. The sentence does not appear on the screen.

The CRM corpus has been used in both energetic and informational masking paradigms (Brungart, 2001a). These phenomena are described using two simultaneous audio streams that may interfere with each other. The masking effect occurs on at least two levels. First, EM happens when two competing streams of interest overlap in the periphery of the auditory system (Durlach et al., 2003). Second, IM refers to the residual interference that remains after accounting for EM (Cooke et al., 2008). This form of masking interacts with EM, and is worsened when the streams are semantically similar, as this creates interference that increases the listener's cognitive load during processing (Johnsrude and Rodd, 2016).

SI was quantified as the proportion of correct responses (in %). A trial was scored as correct when both the color and the digit were accurately retrieved. The resulting percentages were transformed to Rational Arcsin Units (RAU).

The CRM task was programmed in MATLAB and executed on a computer equipped

	English	French
Call signs	Arrow, Baron, Charlie, Eagle, Hopper, Laker, Ringo, Tiger	Alpha, Delta, Charlie, Echo, Kilo, Oscar, Tango, Whisky
Colors	Blue, Green, Red, White	Bleu, Rouge, Vert, Jaune
Digits	One, Two, Three, Four, Five, Six, Seven, Eight	Un, Deux, Trois, Quatre, Cinq, Six, Sept, Huit
Sentences	"Ready [*call sign*], go to [*color*] [*digit*] now."	"Prêt [*call sign*], va au point [*color*] [*digit*] go."

Table 5.1: The CRM variables in the English and French versions of the corpus.

with an RME FireFace UCX sound card[1] to maintain precise control over the auditory parameters.

CRM Words onsets

The onsets of call sign, colors, and digits in the CRM corpus were extracted using a word onset detection script implemented in Python using the free OpenAI Whisper[2] speech recognition model (Radford et al., 2022) in combination with the Whisper timestamp library[3].

1.2 Listening Effort - ESCU

Subjective LE was assessed using the Categorical Listening Effort Scaling method (Luts et al., 2010), which quantifies perceived effort in Effort Scale Categorical Units (ESCU) (Rennies et al., 2014). This scale is composed of seven primary listening effort levels, and six intermediate levels, making a total of 13 possible answers ranging from "*No effort*" (score = 1) to "*Extreme effort*" (score = 13). An additional level "*Only noise*" (score = 14) was also included. The scale was presented in French to the participants in all studies, using a French translation adapted from Lanzilotti et al. (2022) (Figure 5.2).

1.3 Audiometry

Participants' hearing levels were measured using pure-tone audiometry with an Elios® clinical audiometer (Echodia, Le Mazet-Saint-Voy, France; see Figure 5.3). Thresholds were measured at the following frequencies: 0.25, 0.5, 1, 2, 4, 6, 8 and 12.5 kHz, first in the left ear, then in the right. For inclusion, the mean auditory threshold across frequencies was considered. Participants with a mean auditory threshold above 20 dB were excluded from the studies.

[1] https://www.rme-audio.de/fireface-ucx.html
[2] https://github.com/openai/whisper
[3] https://github.com/linto-ai/whisper-timestamped

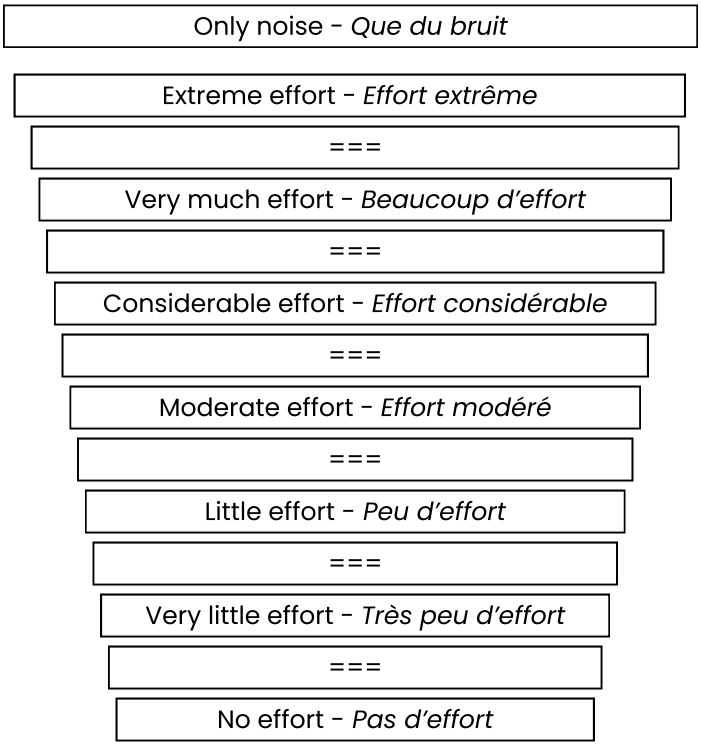


Figure 5.2: Subjective listening effort scale (ESCU) in its French and English versions.

1.4 Speech to text - Stroop task

In this project, the Stroop task relied on recorded verbal responses. Given the high number of trials and sessions, automated transcription was necessary. Therefore, we used the OpenAI Whisper[4] speech recognition model (Radford et al., 2022) to extract the verbal responses of each trial, as well as the response time needed to compute inhibitory cost.

1.5 Resting-State

The resting-state EEG activity was recorded at the start of all EEG sessions of this project. During this 5-minute rest, participants were instructed to alternate between 30 seconds with open eyes and 30 seconds with closed eyes, and to let their thoughts flow while avoiding falling asleep. After the five minutes, they answered the Amsterdam Resting-State Question (ArsQ, see Appendix (ArsQ, Diaz et al., 2013) to report their feelings and experiences during the resting-state.

2 Participants recruitment

Recruiting participants is a complex, but essential, step in behavioral and cognitive studies, as data cannot be collected without them. For this project, a total of 141 different participants were recruited across the three studies. Recruitment

[4] https://github.com/openai/whisper

Figure 5.3: Echodia Elios device with answer button and headset.

employed multiple strategies, including posters, email newsletters, radio, journals, and television interventions. This broad recruitment approach aimed to minimize bias associated with recruiting participants solely from the nearby environment, an aeronautic engineering school, which would likely result in a participant pool predominantly composed of men around 20 years old and introduce potential experiment bias. Overall, the final samples achieved a balanced distribution of men and women across all three studies.

3 Electroencephalography

For this project, the same 64-electrode EEG montage was used across all studies. This quantity of electrode configuration provides sufficient spatial resolution for analyses such as MSs (Section 2.4 in Chapter 2) and source localization (Section 2.4 in Chapter 2).

Different types of EEG analyses can be applied depending on the research question. Here, we describe the pipelines for the approach used in this project for the interpretation of EEG signals in complex auditory situations.

EEG preprocessing, processing, and statistics were performed using Matlab and the EEGlab toolbox (Delorme and Makeig, 2004). EEG analyses focused on the SI task, as this project concerns the neural markers of SI and LE. However, EEG data from the EF tasks could be analyzed in future work. Microstates (MS) analyses were conducted on resting-state recordings acquired at the start of each session, as MS are typically studied during resting-state.

3.1 Frequency, time-Frequency and component Analysis

Preprocessing

The raw EEG data for frequency, time-frequency, and component analyses were preprocessed following the pipeline shown in Figure 5.4. First, channels were selected and relocated, and an average re-reference was computed. Low- and high-pass filters (4-15 Hz; see Box 2.16) were applied, and the data were resampled at 512 Hz. Bad channels were then removed and interpolated to clean the data. For the SI task using the CRM corpus, word onsets (see Section 1.1) were added as EEG triggers for each trial.

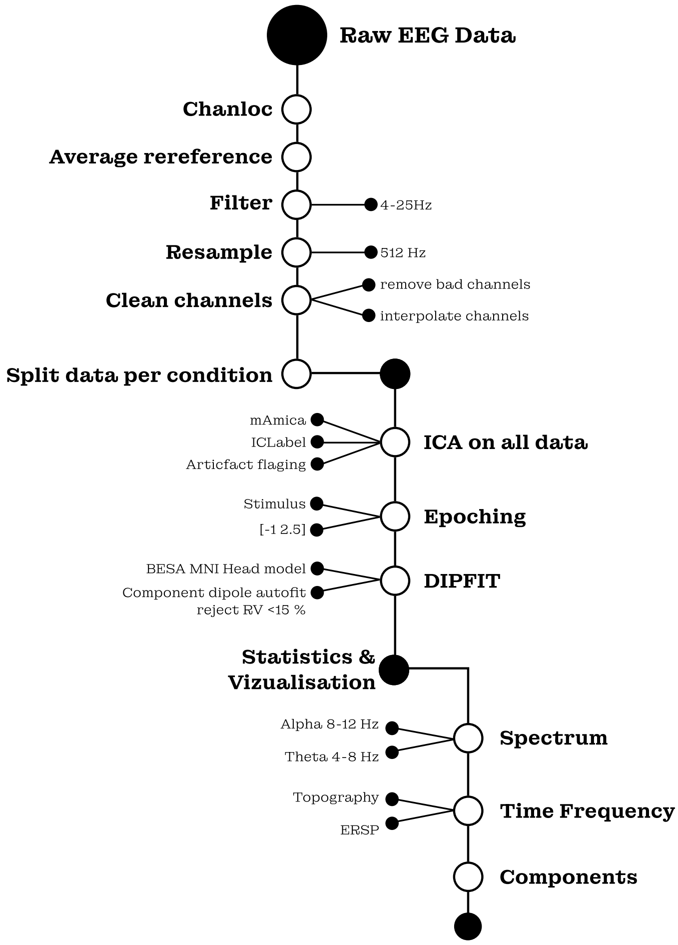


Figure 5.4: EEG preprocessing pipeline for the frequency and time-frequency domain analyses

After this first step of preprocessing, each dataset was split into new datasets corresponding to each condition (SNR or TMR). For each condition dataset, a multimodal ICA was applied using three models (mAMICA[5]) for artifact rejection and subsequent component analyses.

Following preprocessing, the EEG signal was segmented into epochs around each CRM stimulus, using a time window of -1.5 to 2 seconds; each epoch corresponding to a single trial. In addition, the potential EEG sources of components were estimated using the Dipfit EEGlab plugin[6]. For analyses, the sources were located using the Loreta-key Software (Pascual-Marqui et al., 1994).

The filters and mAmica algorithm were chosen based on recommendations of Wisniewski et al. (2024) for the time-frequency and component analyses.

[5]https://github.com/sccn/amica
[6]https://github.com/sccn/dipfit

EEG Study Analyses and Statistics

The preprocessed datasets were used to create studies for visualization and statistical analyses. Independent studies were created for SIS and SIN, containing all TMR or SNR, as illustrated in Figure 5.5. Figures were exported from EEGlab and refined using custom scripts. Statistical analyses were performed using the EEGlab statistics, using permutation, with FDR correction and significance threshold set at 0.05. To improve perceptual accuracy, the parula colormap was used instead of the default jet colormap.

Source Localization

The Loreta-Key software was used to compute the activation sources of the data set extracted from EEGlab.

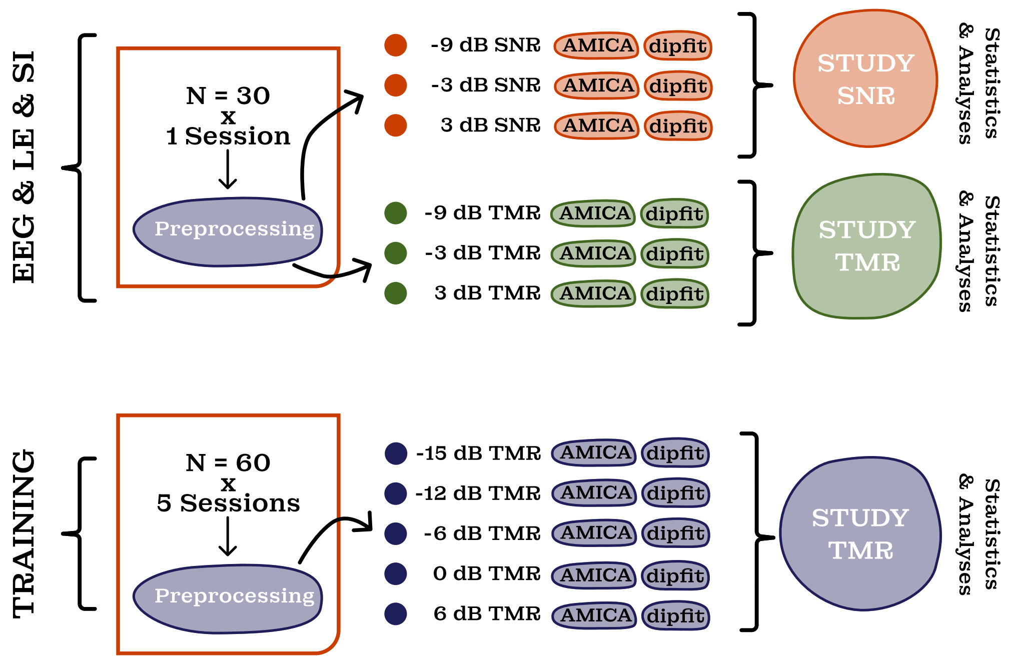


Figure 5.5: Summary of the EEG analysis pipelines for EEG studies of this project.

3.2 Microstates Analysis

MSs are defined in the Section 2.4. Here, we describe the classical analysis pipeline and the additions proposed for this project.

Preprocessing

Each 5-minute resting-state was extracted from each participant's raw EEG data. A re-referencing is applied using the Cz electrode. Using the Automagic toolbox (Pedroni et al., 2019), bad channels were then removed, high-pass (1 Hz) and notch (50 Hz) filters were applied, as well as an average re-reference. ICA and interpolation, using the default Automagic parameters, were then applied to remove components related to eye movements, muscle activity, and cardiac artifacts.

After preprocessing (see Figure 5.6), only the closed-eyes segments (half of the resting-state recording) were retained. The data were then epoched into 2-second segments, and epochs with amplitudes exceeding a 90 µV threshold were rejected.

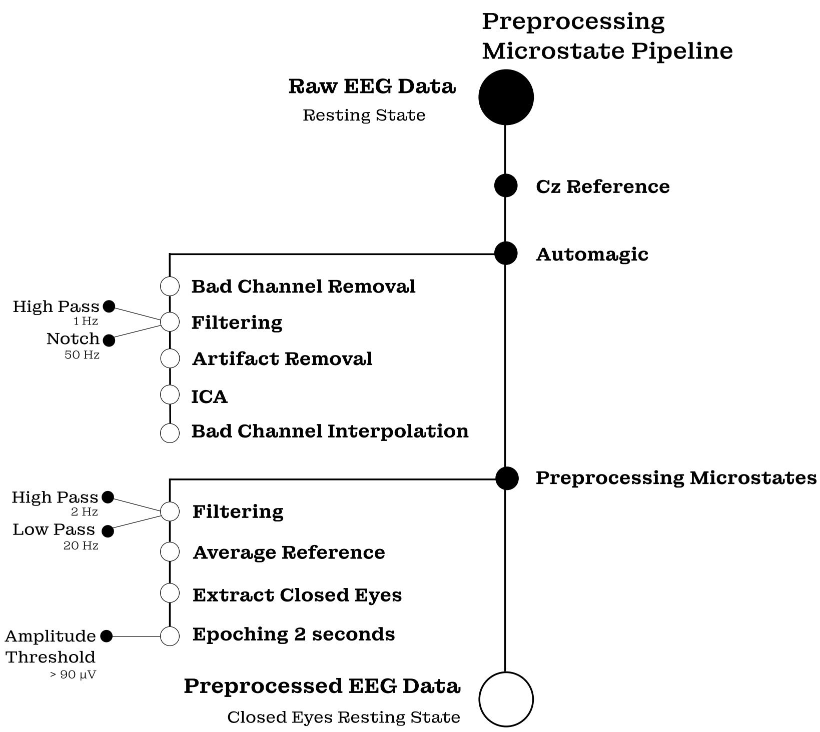


Figure 5.6: Microstates preprocessing pipeline.

Microstates classic pipeline

The classical pipeline is illustrated in Figure 5.7. After data preprocessing, topographical maps of each GFP peak are extracted for each data set. These maps are then clustered using the k-means algorithm (see Box 3.2).

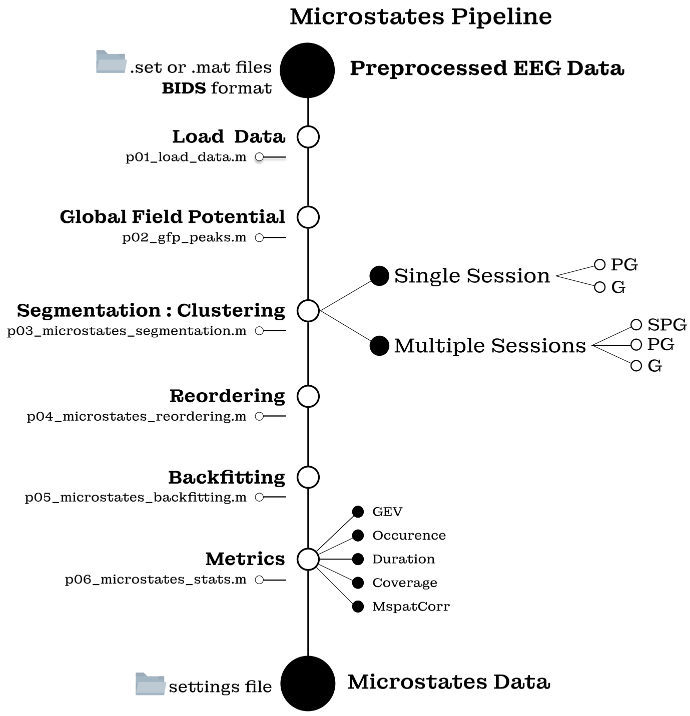


Figure 5.7: Microstates Analysis Pipeline

K-mean Clustering

K-means clustering is an unsupervised machine learning algorithm that uses the concept of nearest neighbors to create clusters in single or multi-dimensional data. The user specifies the desired number of clusters (k) as input. At first, k cluster centroids are randomly selected. Each data point is then assigned to the nearest centroid based on distance calculation. Then, the mean position of each cluster is computed and set as the new centroid. This process is iteratively repeated for a predefined amount of times or until the clusters stabilize. Since the initial centroids are randomly selected, the entire clustering loop is repeated multiple times to minimize variability. The best clustering fit of all iterations is then retained as the final clustering results. For the MSs analysis, the mean map of each cluster serves as the prototypical map used for backfitting.

The number of clusters (k) is predefined based on literature or the global explained variance (GEV). Canonically, four clusters are used; however, the methodology may vary depending on the authors. The number of clusters can be determined by analyzing the reduction in variation relative to k, often visualized with an elbow plot. In MSs analysis, the number of clusters corresponds to the number of MSs in the prototypical maps. While computing clusters, multiple k values can be tested, and their Global Explained Variance (GEV) can be compared to select the optimal number of MSs.

The K-means clustering produces k prototypical maps representing the group.

These group-representative maps, called prototypical MSs, are labeled according to the standard nomenclature using the Latin alphabet. The backfitting is performed, during which each GFP peak map is labeled according to the less distant prototypical group maps. This procedure results in a sequence of labeled MSs (Figure 5.8). Metrics such as occurrence, duration, coverage, and global explained variance (GEV) of each MS (see Section 2.4) can then be extracted from this sequence.

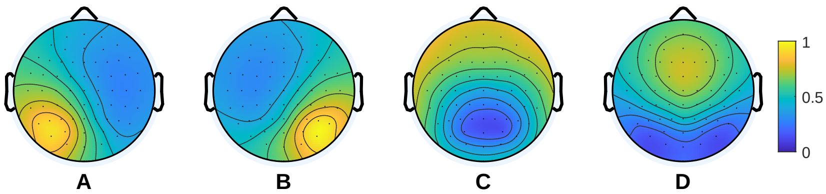


Figure 5.8: Example of microstate output (k=4)

Addition to the classic pipeline

In the literature, MSs analysis is often applied following the classical pipeline, where prototypical maps are computed across the entire dataset, as a group. This approach is sufficient for simple experimental protocols involving a single session or a single group. However, for more complex studies or for clinical design, it may be advantageous to extend the MS analysis pipeline.

For protocols with multiple sessions per participant, computing group-level prototypical MS maps and backfitting is informative but overlooks the contributions of individual subjects and sessions. While MSs are highly reproducible both within and between individuals, their metrics also exhibit variability.

The original pipeline was extended to compute MS at the session, participant, and group levels, increasing the granularity of information.

To implement this, we developed a modified MS analysis pipeline (see Figure 5.9) that can be adapted depending on protocol-specific, such as repeated measures (multiple sessions) or the need for group-level analyses. This pipeline was implemented not only for the present project but also for other studies, including clinical applications (Lebely et al., 2024).

In a multi-session experimental design with S sessions and P participants, the pipeline can follow a session-participant-group structure (see Figure 5.9). The classical pipeline is first applied at the session level, where k-means clustering is performed on each GFP peak, producing S x P prototypical maps (for each session of each participant). Next, clustering is applied at the participant level on the MS of the previous level (the session level), resulting in P prototypical maps (for each participant). At last, clustering is performed at the group level on the MS of the previous level (the participant level), resulting in one set of prototypical maps. Each set of prototypical maps is composed of k cluster centroids.

Backfitting is performed according to user-defined parameters, either at all levels or selected levels. Users may also skip levels; for example, concatenating session or participant datasets allows the pipeline to run in a participant–group or group-only design. As in the classical pipeline, the user can select the desired number of clusters k.

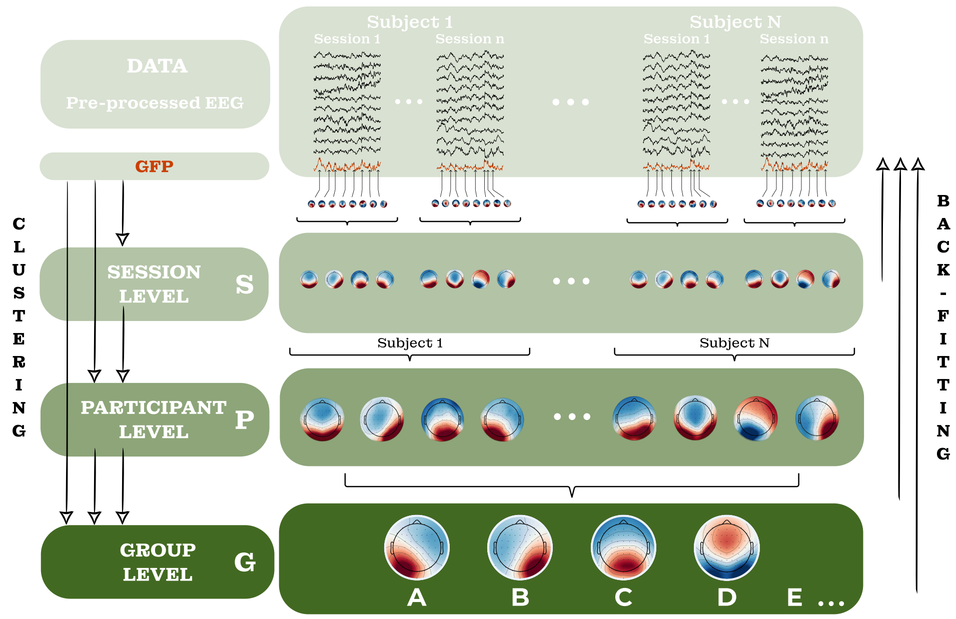


Figure 5.9: Adapted MS pipeline for session, participant, and group levels MS analyses.

The scripts for the Matlab GUI and MSs analysis can be found on the following repository: https://github.com/neuroergoISAE/EEGMicrostates.git.

Publication

The toolbox implemented in Matlab was presented in a poster at the Microstates Conference in 2024 in Geneva (see Appendix **??**).

4 Data

4.1 Data recording

Electrophysiological and behavioral data were recorded simultaneously and synchronized using Lab Streaming Layer[7] (LSL). EEG data were saved in *.xdf* format for each session independently. All data were organized according to BIDS standards to improve usability.

4.2 Data size

Working with large numbers of participants and multiple sessions results in rapidly increasing data volumes. EEG signals were resampled at 512 Hz before saving to reduce file size. In addition, parallel computing was employed to decrease processing time, using the MATLAB Parallel Computing Toolbox for EEG analyses and the Python multiprocessing package for speech-to-text processing.

[7]https://github.com/sccn/labstreaminglayer

4.3 Repository and Data availability

Data are available on demand. All scripts used for the experimental procedure and analyses are available on GitHub repositories as detailed in the following list.

Microstates https://github.com/neuroergoISAE/EEGMicrostates.git

Resting State https://github.com/neuroergoISAE/RestingState.git

First study https://github.com/ipiup/CRMFREN.git

Second study https://github.com/ipiup/CRMEF.git

Third study https://github.com/ipiup/CRMEF_Training.git

6

Language

Description of the Chapter

The demographics of participants included in studies largely depend on the population surrounding the laboratory. In universities, many young adults usually have access to information about participant recruitment, whereas hospitals tend to involve older populations. The language spoken and understood by participants also depends on the laboratory's location. Despite English being the primary language of scientific communication, many research facilities are located in non-English-speaking countries. Therefore, participants performing experimental tasks may not be fluent in English, and tasks must be adapted accordingly, especially in speech research. This chapter presents the first study of the project, which aimed to assess differences in speech understanding in complex auditory conditions of native French listeners using the Coordinate Response Measure in its original English version and in its French version.

Hypothesis 1

The use of a native language influences the intelligibility and LE, particularly in adverse situations.

Main Results

SI improved and LE decreased with increasing SNR/TMR with both native and non-native language corpora.

Non-native language affected SI and LE differently than the native language corpus.

Energetic masking affected SI and LE strongly in a non-native language.

Informational masking affected SI and strongly in a non-native language, especially in adverse conditions.

A native-language version of the corpus is recommended for further studies.

1 Research Goal

A listener's capacity to understand speech in complex auditory environment depends not only on their sensory acuity, but also on cognitive factors such as selective attention to a target talker or foreign language proficiency. These factors interact and require the listener to allocate mental resources to understand the target speaker, a demand that has been described in the literature as LE (Peelle, 2018; Pichora-Fuller et al., 2016). Furthermore, in complex auditory situations, it becomes even more challenging to understand a second language compared to one's native language (Cooke et al., 2008; Rogers et al., 2006; Warzybok, Brand, et al., 2015). While this phenomenon can be investigated experimentally, using second-language stimuli to assess SI may introduce measurable and avoidable biases. In particular, the original English language version of the CRM corpus is predominantly used to assess both native and non-native English listeners. This population heterogeneity can result in a wide variation in comprehension abilities, potentially introducing performance bias and undermining reproducibility. For instance, research has reported that a noisy environment affects non-native listeners more (Mattys et al., 2012) and increases their LE in SIS tasks (Cooke et al., 2008). In addition, interference between two English audio streams may impair SI differently compared to interference between streams in the listeners' native language.

In this first study, we hypothesized that the cognitive load associated with SI increases when listening in a second language, thereby increasing the LE exerted by the listener. The aim was to quantify the impact of using non-native language stimuli in order to assess bias in experiments designed to evaluate multi-talker listening situations.

The CRM corpus enables the measurement of a dependent variable, specifically the percentage of correct responses, which is influenced by independent variables such as TMR or SNR. When a non-native language corpus is used, language proficiency is another co-factor. Several versions of the CRM corpus have been developed in languages such as Spanish (Lelo de Larrea-Mancera et al., 2023), Dutch (Nagels et al., 2021), Persian (Amiri et al., 2020), British English (Kitterick et al., 2010; Semeraro et al., 2017), Kannada (Rachana and Neelamegarajan, 2024) and Mandarin (Wang et al., 2019). A French version was also recently developed by Isnard et al. (2024).

These localized language versions of the CRM corpus make it possible to eliminate language as a confounding factor and enable a more accurate comparison of the core independent variables. Alternatively, language can be included as an experimental variable in assessments of the additional LE associated with processing speech in a non-native language. Therefore, this first study aims to determine the impact of using English and French versions of the CRM corpus on SI and LE in native French speakers in both SIN and SIS conditions.

This first study seeks to address the first hypothesis of the project and to guide the choice of language for the CRM for the next stages of the project.

Hypothesis 1

The use of a native language influences the intelligibility and LE, particularly in adverse situations.

2 Material and Methods

2.1 Participants

Fifty-one participants were recruited for this experiment (29 women, 1 other, mean age of 26.4 ± 5.4 years). All had normal hearing, as confirmed by pure-tone audiometry using an Elios® clinical audiometer (Echodia, Le Mazet-Saint-Voy, France). Hearing thresholds were assessed at the following frequencies: .25, .5, 1, 2, 4, 6, 8 kHz and 12.5 kHz (hearing level $\leq$ 20 dB HL; mean hearing level -.3 ± 8.3 dB HL, details in Appendix 10.1).

To be eligible, participants had to meet the following criteria: no known hearing impairments, aged between 18 and 40 years, native speaker of French, no uncorrected visual impairments, and adequate English comprehension, subjectively self-assessed using the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007) (see Table 6.1). Participants received €30 in financial compensation for their involvement in the study.

Measure	**Mean**	**SD**
Oral production	6.45	1.65
Oral comprehension	7.04	1.70
Reading	7.51	1.46
LEAP-Q Mean	7.01	1.49

Table 6.1: Mean and standard deviation (SD) for subjective assessment of English Proficiency. Oral production, Oral comprehension and Reading were assessed on a scale of 1 to 10. LEAP-Q mean represents the mean of the three subjective values.

2.2 Ethics Approval

This study was approved by the local ethics committee (IRB Number 2023 647). All participants provided written informed consent prior to data collection.

2.3 Stimuli

Stimuli were retrieved from both English and French versions of the CRM corpus. In the original English form, a typical sentence follows a fixed structure: "Ready [call sign], go to [color] [digit] now", where call sign, color, and digit are selected from a predefined word list (Bolia et al., 2000) (Table 5.1). The corpus includes eight call signs, four colors, and eight digits, resulting in 256 combinations. These were recorded by eight talkers (four men and four women), making a total of 2048 English-language stimuli. The French language version of the CRM has been designed to mirror, as closely as possible, the English language version (Isnard et al., 2024). The sentence structure remains almost identical: "*Prêt* [call sign], *va au point* [color] [number] *go*" with equivalent lexical variables (see Table 5.1 in Chapter 5). As with the English language version, eight French talkers (four men and four women) recorded the 256 sentence combinations, resulting in 2048 French-language stimuli. In both languages, the recorded sentences range in duration from 2 to 3 seconds.

In this study, the two versions of the CRM corpus were used in three distinct experimental conditions:

SIN condition: In this condition, a single target talker was presented alongside background noise. The SNR corresponded to the ratio of the sound level of the target talker (signal) to the background (noise). The noise was generated as Gaussian noise, spectrally shaped using a finite impulse response (FIR) filter to match the spectral profile of sentences in the CRM corpus, as described in Brungart (2001b).

SIS condition: Two talkers were presented simultaneously, with the TMR corresponding to the difference between the sound level of the target and the masker. Talker pairings consisted of either 1 woman and 1 man, 2 women, 2 men, or the same talker. The two simultaneous sentences always featured different call signs and different color/digit combinations.

Control condition: A target talker was presented without any masking signal (i.e., in silence), in both languages, to confirm that the participant was able to understand speech clearly in the absence of interference.

2.4 Apparatus

Participants were seated in a quiet experimental room, facing a screen that displayed the response button options used in the experimental tasks. Graphical user interfaces were developed in MATLAB R2021a (MathWorks). First, a matrix of 32 buttons was displayed. Buttons were organized into four rows (one for each color) and eight columns (one for each digit) for the SI task (Figure 5.2). Second, subjective LE was rated using the Categorical LE Scaling method (Luts et al., 2010), as described in Section 1.2 in Chapter 5

Auditory stimuli were presented diotically via Beyerdynamic DT-770 headphones, connected to an RME Fireface UCX sound card. Stimuli were delivered at approximately 55 dB SPL, with a sampling rate of 44.1 kHz.

2.5 Procedure

After providing informed consent, participants completed pure-tone audiometry. They then completed a demographic questionnaire and the LEAP-Q (Marian et al., 2007). The LEAP-Q enabled us to assess each participant's English comprehension based on measures of language exposure, experience, and proficiency (see Table 6.1). Participants subjectively rated their English proficiency on a scale ranging from 1 to 10 for oral, listening, and reading skills. The average of these three ratings was used as a measure of their subjective proficiency.

Participants then completed the SI task. In each experimental block, they were asked to indicate the color and digit associated with the call sign "*Baron*" (in English sentences) or "*Delta*" (in French sentences), by clicking the corresponding button. There was no time limit for responses. The order of presentation of the experimental conditions was counterbalanced, using either the English or French version of the CRM, and either the SIN or the SIS masking condition, according to a predefined Latin square design. For the SIN condition, nine SNR levels were tested (ranging from -18 dB to +6 dB, in 3 dB steps). For the SIS condition, nine TMR levels were used (ranging from -12 dB to +12 dB, also in 3 dB steps).

For each masking level, participants completed 32 trials, representing all possible combinations of 4 colors x 8 digits, selected from recordings by the 8 different talkers. In each trial, the talker(s) were randomly chosen to minimize talker-specific effects.

Finally, after each block, participants were asked to rate their subjective LE. A control block, consisting of 32 trials, was presented after both SIS and SIN conditions, in both French and English versions. Here, the aim was to ensure that participants clearly understood the stimuli under neutral (no masker) conditions. Each participant completed a total of 38 blocks, each containing 32 trials: 18 SIS blocks (32 trials x 9 TMR levels x 2 languages) + 18 SIN blocks (32 trials x 9 SNR levels x 2 languages) + 2 control blocks (32 unmasked trials x 2 languages). Each experimental block lasted approximately 2 to 3 minutes, and participants were allowed to take short breaks between blocks. The full experimental session lasted approximately 2 hours.

2.6 Data analyses

The percentage of correct responses was calculated as the number of trials in which both the color and digit were correctly identified, divided by the total number of trials in that experimental condition (32 trials in each block). Following current recommendations for speech analyses, scores were normalized using the Rationalized Arcsine Unit (RAU) transform in each condition (Studebaker, 1985).

Statistical analyses were carried out using MATLAB and R (version 4.0.3). For both SIN and SIS conditions, a two-way analysis of variance (ANOVA) with Greenhouse–Geisser correction for sphericity was conducted to examine the main effect of CRM language (English vs. French) and masking level (9 SNRs or 9 TMRs), as well as their interaction. The significance threshold was set at *p* ¡ .05. Tukey's HSD (Honestly Significant Difference) post-hoc test was applied for pairwise comparisons when the effect was significant.

3 Results

For the sake of clarity, SI results in plots are presented as the percentage of correct responses; however, all statistical analyses were conducted on RAU-transformed data.

3.1 Speech-in-Noise

SI

In the SIN condition, the ANOVA showed a significant main effect of SNR ($F(8, 400) = 2747.81, p < .001, \eta^2 = .982$; see Figure 6.1 and Table 6.2). A significant main effect of CRM language was also observed ($F(1, 50) = 1199.77, p < .001, \eta^2 = .960$), with better intelligibility in French compared to English. Additionally, a significant interaction between language and SNR level was found ($F(8, 400) = 115.77, p < .001, \eta^2 = .698$). Tukey's HSD post-hoc test showed that participants performed significantly better in French than in English for negative and neutral SNR levels (i.e., when the masker level was equal to or higher than the target level), except at the lowest SNR (-18 dB).

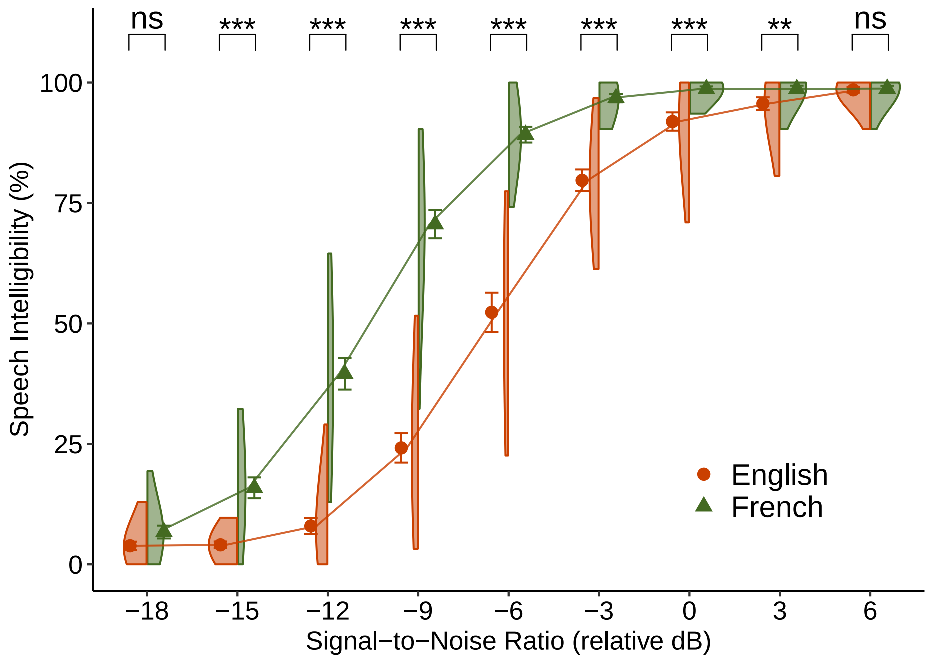


Figure 6.1: Mean SI (percentage correct) in SIN condition, for each SNR level (-18 dB to +6 dB SNR; in 3 dB steps). Vertical bars indicate the confidence interval at 95%. Violin plots represent response distribution density. Statistical comparisons were performed on RAU-transformed data. Post hoc significance: *: $p < .05$, **: $p < .01$, ***: $p < .001$, ns: not significant.

Subjective LE

An analysis of ESCU responses after each of the 9 SIN blocks showed a significant main effect of language ($F(1, 50) = 262.81, p < .001, \eta^2 = .840$; see Figure 6.2 and Table 6.2), with participants reporting greater subjective LE in English than in French. Additionally, there was a significant main effect of SNR level on subjective LE ($F(8, 400) = 844.52, p < .001, \eta^2 = .944$); specifically, subjective effort increased as SNR decreased. Furthermore, a significant interaction was observed between language and SNR level ($F(8, 400) = 22.33, p < .001, \eta^2 = .309$). Tukey's HSD post-hoc test showed that participants reported significantly greater LE in English compared to French across all SNR levels, except in the two most adverse conditions (-18 dB and -15 dB SNR).

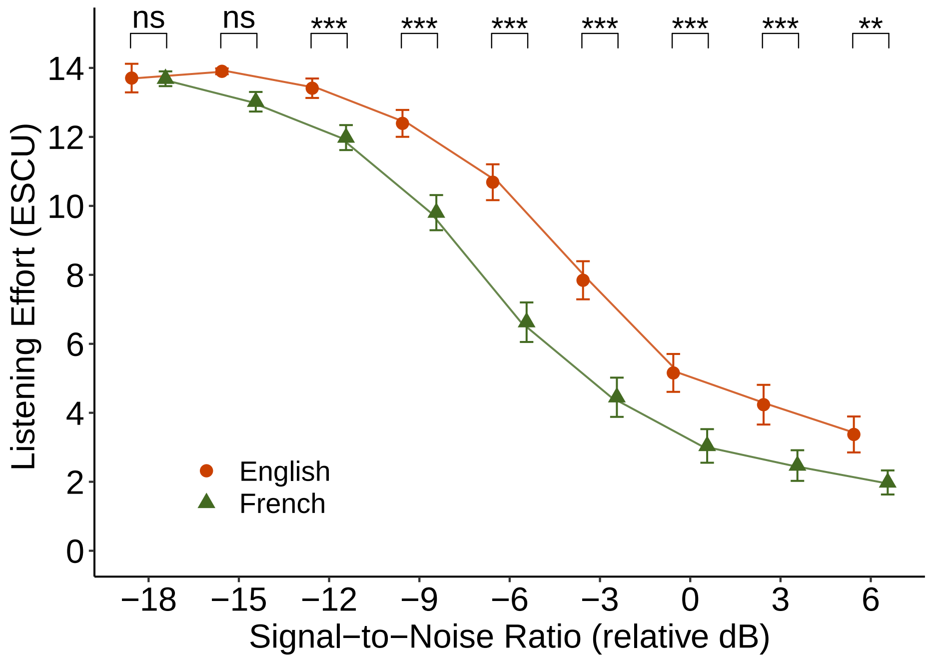


Figure 6.2: Mean LE (ESCU) in SIN condition, for each SNR level (-18 dB to +6 dB SNR; in 3 dB steps). Vertical bars indicate the confidence interval at 95%. Statistical comparisons were performed on RAU-transformed data. Post hoc significance: *: $p < .05$, **: $p < .01$, ***: $p < .001$, ns: not significant.

		SI				LE			
		Df	F	η_p^2	p-value	Df	F	η_p^2	p-value
SIN	SNR	8	2600.42	.981	$<$ **.001**	8	844.52	.944	$<$ **.001**
	Language	1	1048.25	.954	$<$ **.001**	1	262.81	.840	$<$ **.001**
	Language:SNR	8	91.66	647	$<$ **.001**	8	22.33	.309	$<$ **.001**
SIS	TMR	8	293.17	.854	$<$ **.001**	8	259.58	.838	$<$ **.001**
	Language	1	246.90	.832	$<$ **.001**	1	25.71	.340	$<$ **.001**
	Language:TMR	8	23.98	.324	$<$ **.001**	8	8.13	.140	$<$ **.001**

Df: degrees of freedom, F: F-test value, η_p^2: effect size

Table 6.2: Main ANOVA results for SI (CRM) and LE (ESCU) in SIS and SIN conditions.

3.2 Speech-in-Speech

SI

In the SIS condition, the two-way ANOVA showed a significant main effect of TMR level ($F(8, 210) = 268.173, p < .001, \eta^2 = .843$; see Figure 6.3 and Table 6.2). Furthermore, a significant main effect of language was observed ($F(1, 50) = 270.717, p < .001, \eta^2 == .844$), with better intelligibility in French compared to English. A significant interaction between language and TMR level was also found ($F(8,400)\eta^2 = 29.531, p < .001, = .371$). Post-hoc tests revealed that this interaction was due to participants performing better in French than in English at negative and neutral TMR levels (i.e., when the masker was equal to or louder than the target signal).

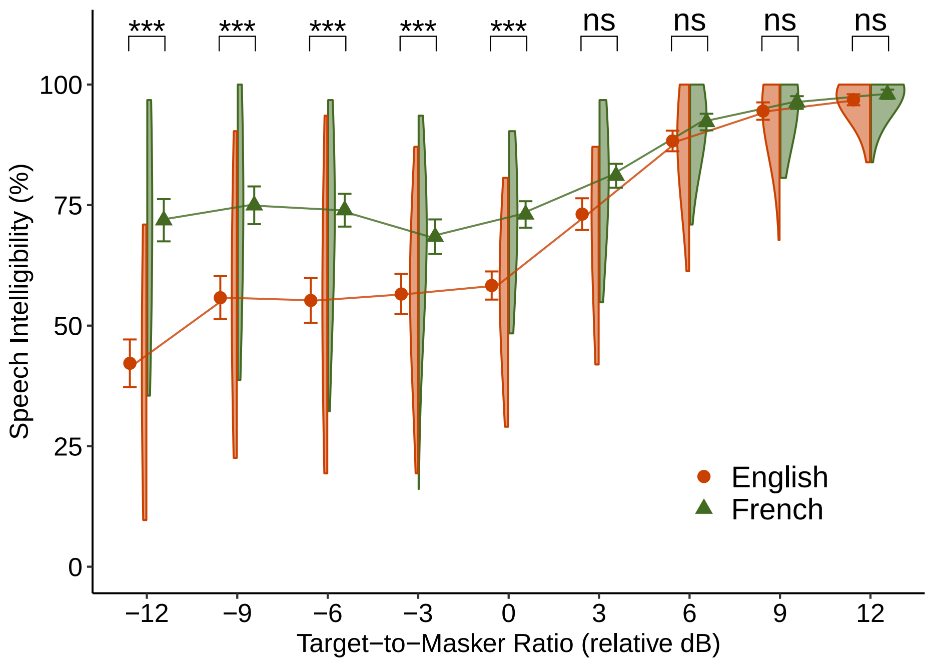


Figure 6.3: Mean SI (percentage correct) in SIS condition, for each TMR level (-18 dB to +6 dB SNR; in 3 dB steps). Vertical bars indicate the confidence interval at 95%. Violin plots represent response distribution density. Statistical comparisons were performed on RAU-transformed data. Post hoc significance: *: $p < .05$, **: $p < .01$, ***: $p < .001$, ns: not significant.

Subjective LE

A two-way ANOVA of ESCU ratings after each of the 9 SIS blocks showed a significant main effect of language ($F(1, 50) = 25.71, p < .001, \eta^2 = .340$; see Figure 6.4 and Table 6.2), with participants reporting greater subjective effort in English compared to French. Additionally, a significant main effect of TMR level on subjective LE was observed ($F(8, 400) = 259.58, p < .001, \eta^2 = .838$). A significant interaction between CRM language and TMR level was found ($F(8, 400) = 8.13, p < .001, \eta^2 = .140$). Tukey's HSD post-hoc test showed that participants reported significantly greater LE in English compared to French across nearly all TMR levels, except in the two most adverse conditions (-12 dB and -6 dB TMR).

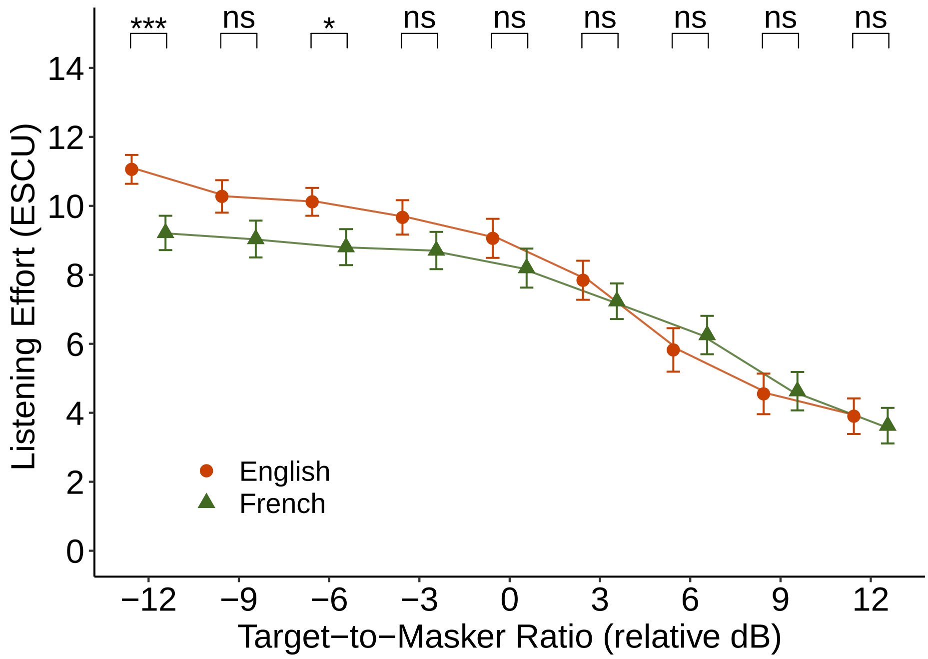


Figure 6.4: Mean LE (ESCU) in SIS condition, for each SNR level (-18 dB to +6 dB TMR; in 3 dB steps). Vertical bars indicate the confidence interval at 95%. Statistical comparisons were performed on RAU-transformed data. Post hoc significance: *: $p < .05$, **: $p < .01$, ***: $p < .001$, ns: not significant.

3.3 Post-hoc and supplementary figures

The Figure 6.5 shows the post-hoc results between each SNR in SIN and each TMR in SIS.

a. SIN **b. SIS**

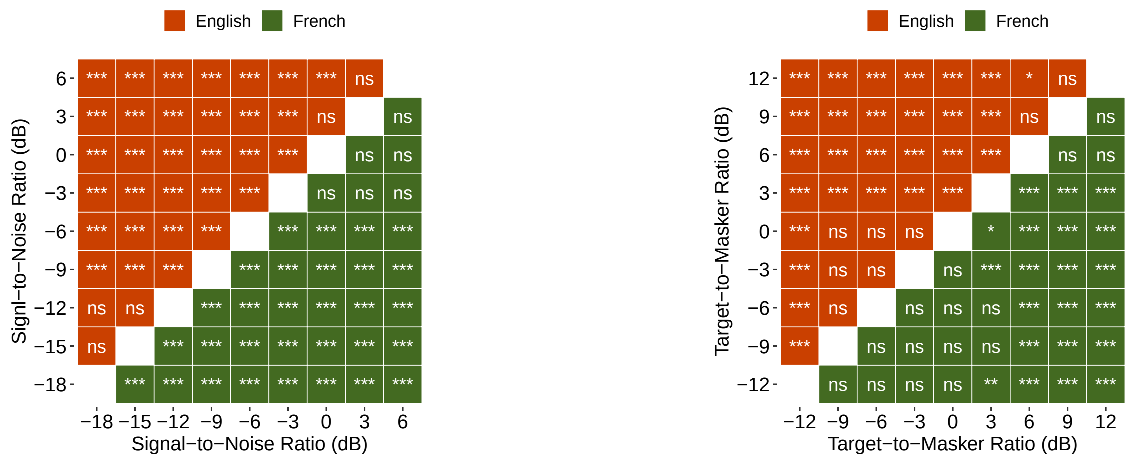


Figure 6.5: Speech intelligibility comparison between SIN and SIS. Tukey's HSD p-values in SIN condition, for each SNR. b. Tukey's HSD p-values in SIS condition, for each TMR. Post hoc tests were performed on RAU-transformed data. Significance levels: * :$p < .05$, **: $p < .01$, ***:$p < .001$, ns: not significant.

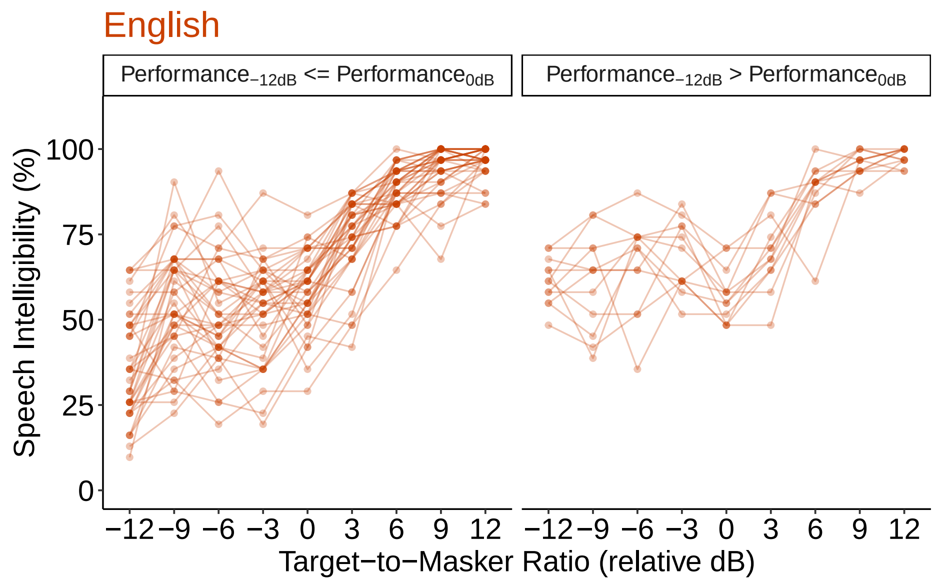


Figure 6.6: Performance of individual participants. SI (%) at each TMR level in the English condition. Left panel: Participants whose performance at -12 dB TMR was greater than, or similar, to their performance at 0 dB TMR. Right panel: Participants whose performance at -12 dB TMR was lower compared to 0 dB TMR.

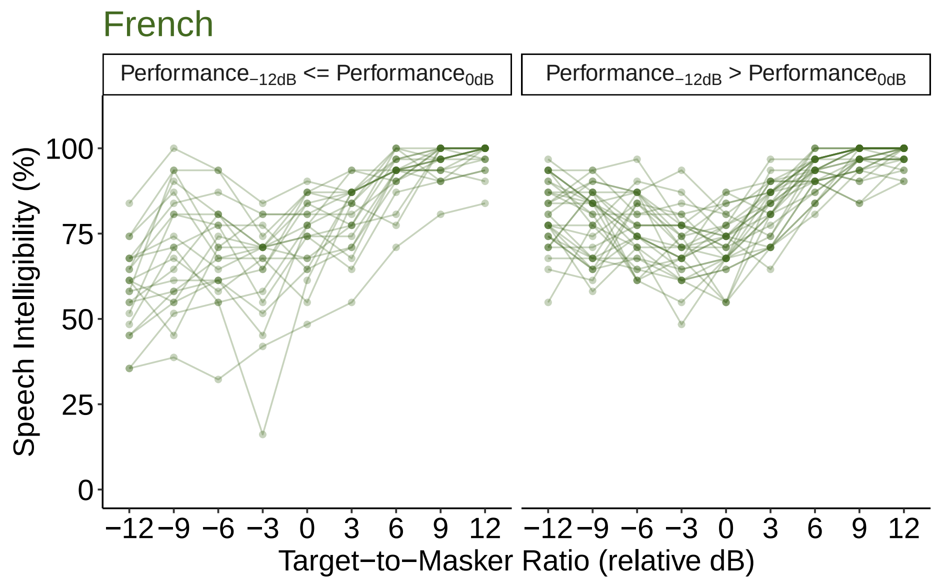


Figure 6.7: Performance of individual participants. SI (%) at each TMR level in the French condition. Left panel: Participants whose performance at -12 dB TMR was greater than, or similar to their performance at 0 dB TMR. Right panel: Participants whose performance at -12 dB TMR was lower compared to 0 dB TMR.

3.4 Type of response

To investigate whether incorrect responses in the SIS condition were more influenced by masker interference than random guessing, we analyzed participants' responses with respect to digit- and color-related errors. In the majority of incorrect trials, participants preferentially chose information corresponding to the masker rather than making a random decision. This was observed for the color, the digit only, or both. The distribution of response types is illustrated in Figure 6.8.

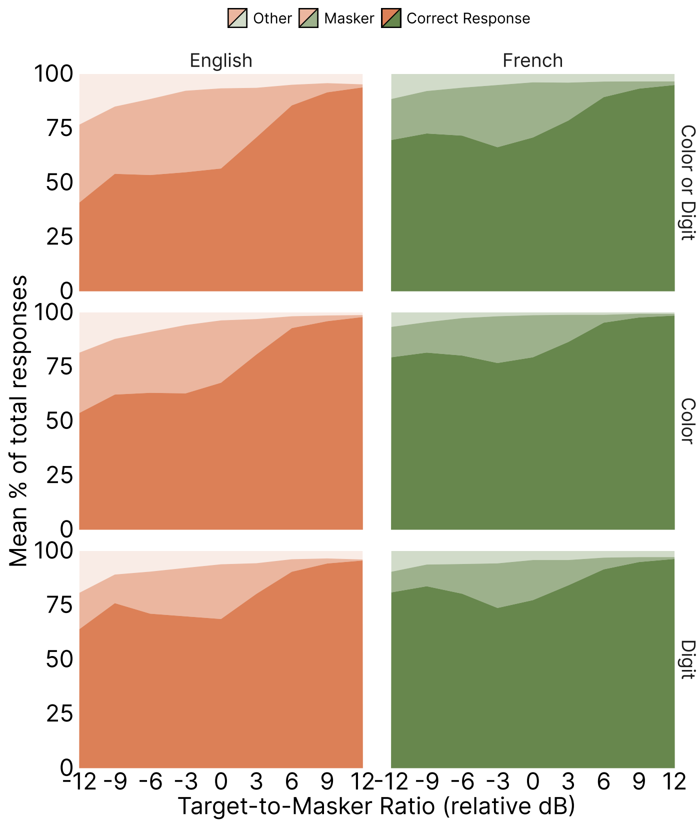


Figure 6.8: Mean percentage of response type for English and French CRM in SIS condition. Correct responses, masker responses, or other responses for color, digit, or both, for each TMR level.

3.5 English proficiency

Subjective English proficiency was assessed using the LEAP-Q. Three variables were selected: oral production, oral comprehension, and reading. Each variable was rated on a scale from 1 to 10, with 10 representing perfect proficiency. The mean of these three ratings was used as the overall proficiency score. The mean LEAP-Q score across participants was 7.0 ± 1.5 (SD). Results are shown in Table 6.1.

This measure confirmed a minimum level of English proficiency in the same way that audiometric testing ensured that participants had no hearing loss.

4 Discussion

The goal of the present study was to evaluate the effects of language (native vs. non-native) on SI and LE.

Hypothesis 1

The use of a native language influences the intelligibility and LE, particularly in adverse situations.

English and French versions of the CRM corpus were used with native French listeners under two conditions: (1) SIN, in which a single talker was masked by stationary speech-shaped noise; and (2) SIS, involving two simultaneous talkers. LE was assessed in each condition using the Categorical LE Scale method (Luts et al., 2010; Rennies et al., 2014).

4.1 Speech-in-Noise

As expected, in the SIN condition, intelligibility gradually declined as SNR decreased. This main effect of SNR is aligned with previous findings reported in the literature (Brungart, 2001a, 2001b). We also found that intelligibility in the presence of background noise was influenced by the language. As expected, French participants performed better when listening in French compared to English. This suggests that the louder the relative masking noise, the more difficult it is for the listener to separate the target from the noise, and that this challenge is greater when the speech is in a non-native language.

These intuitive results provide useful reference values for SIN tests. With the exception of the most adverse condition (-18 dB SNR), where a floor effect was observed, and the most favorable conditions (+3 and +6 dB SNR), where intelligibility reached its maximum, participants consistently performed better in French than in English. This pattern suggests a language-related effect that interacts with the level of energetic masking.

The significant interaction between language and SNR, however, is of particular interest. Our results suggest that energetic masking influences speech processing differently in native and non-native languages. In French, intelligibility differed significantly between the two most adverse SNR levels (-18 and -15 dB SNR; see Figure 6.5a for Tukey's post-hoc results), while this was not the case in English. In addition, a ceiling effect appeared at lower SNR levels in French compared to English (no significant difference between -3 and 0 dB SNR; see Figure 6.5a), showing that participants were able to cope better with louder background noise in their native language.

The absence of a language effect in the most adverse condition can be explained by the high level of masking noise. In this condition, whatever the language, masking noise was loud enough to prevent any understanding of the target speech, demonstrated by very low intelligibility scores. ESCU ratings for this SNR level provide further support for the disengagement or withdrawal of participants during this block; most participants selected the "*only noise*" level.

4.2 Speech-in-Speech

In the SIS condition, the results confirmed our first hypothesis regarding a significant effect of task difficulty (i.e., manipulated by TMR levels) on SI (see Figure 6.3). Intelligibility was better at positive TMRs, where the target voice was louder than the masker. This main effect, which is consistent with previous studies (Brungart, 2001b; Thompson et al., 2015), shows that intelligibility decreases as the masker becomes louder relative to the target. Earlier work has shown that informational masking in competing speech streams interferes with speech processing and makes segregation more difficult (Johnsrude and Rodd, 2016).

In SIS tasks, masking is both energetic and informational. Evidence for informational masking can be found in the nature of participants' errors. A closer examination of response types highlighted that incorrect answers were predominantly biased toward the competing talker. It is known that the more similar the competing voices, the more difficult it is to segregate them (Brouwer et al., 2012). Thus, the errors made by participants could be related to both the TMR and/or the acoustic characteristics of the voices (Brungart, 2001b). In addition, as in the SIN condition, the auditory stream from the masker talker introduces a degree of energetic masking. However, the result reported here, namely that errors align with the masker talker, strongly suggests that the dominant form of interference was informational in the SIS condition.

The effect of task difficulty on intelligibility was observed both in French and English versions of the CRM corpus. However, as in the SIN condition, the effect was more pronounced in English. Participants were better able to segregate the target talker when listening in French compared to English. This is probably because they were all native French listeners. Earlier work has demonstrated that knowledge, familiarity, and expertise minimize informational masking in multi-talker scenarios (Johnsrude et al., 2013). This significant primary effect of language is underlined by large effect sizes in both conditions (SIN: = .982; SIS: = .843).

Overall, mean intelligibility scores for English and French ranged from 2 to 28 RAU, with higher scores in French. In addition, performance in the French language tended to stabilize around 70 RAU for the adverse condition, whereas it rapidly decreased in the English language, reaching a minimum of 40 RAU at -12 dB TMR. This marked drop shows how, in a non-native language, informational masking not only affects intelligibility at a lower TMR, but also that the effect is stronger compared to a native language.

Since the energetic masking produced by the competing talker is similar in both languages, the difference in performance is most likely to be driven by increased informational masking in the non-native language. This finding supports the idea that informational masking has a stronger impact when listeners process speech in a non-native language, particularly at the lowest TMRs (Lecumberri et al., 2010; Van Engen and Bradlow, 2007).

The primary outcome of this study, however, reveals an interaction between the CRM language and the TMR level in the SIS condition. Under adverse conditions (i.e., when the target talker was at the same or lower auditory level than the masker), French listeners were better at segregating the target talker in French compared to English, while intelligibility did not significantly differ in favorable conditions. French listeners performed better in the French SIS condition in the range +3 to -12 dB TMR, and their subjective LE was rated as significantly higher in English compared to French (see Figure 6.4 and 6.5b).

While, in general, the effects of energetic masking decrease monotonically with increasing SNR (Brungart, 2001a; Thompson et al., 2015), the presence of informational masking introduces additional variability and engages more complex cognitive processes (Mattys et al., 2012). Examination of individual intelligibility curves revealed distinct performance patterns, suggesting that participants may have different segregation and inhibition strategies. Some showed a monotonic relation between TMR and intelligibility (i.e., the lower the TMR, the lower the intelligibility, as observed in SIN condition), while for others, intelligibility was minimal at 0 dB TMR, and performance was equal or better for the most adverse TMRs.

The results showed that 30 participants performed equally well or better in the range -12 to 0 dB TMR in the French condition (see Figure 6.7), whereas this pattern was only observed in 12 participants in the English condition (see Figure 6.6). As reported by other authors (Andeol et al., 2011; Lanzilotti et al., 2022; Thompson et al., 2015), these results can be explained by the listener's ability to use the cue of a softer voice to segregate the target talker from the masker. While both native and non-native listeners tend to adopt similar strategies when actively listening in complex situations, beyond a certain level, non-native listeners are more adversely affected by environmental distortion and complexity (Bradlow and Alexander, 2007; Brouwer et al., 2012). In our study, we argue that an additional layer of complexity emerged: listeners seemed to be more adept at leveraging these 'negative level cues' in their native language.

It is also important to note that performance significantly improved between -12 dB and -9 dB TMR in English, whereas no such improvement was observed in French. In this TMR range, the psychometric function displayed floor and ceiling effects, demonstrated by the absence of significant differences in adjacent levels (from -12 to 0 dB, and from +6 to +12 dB). Similarly, we observed significant differences between 0 and +3 dB TMR and between +3 and +6 dB TMR in English. The floor effect, however, appeared to persist beyond the initial decline, as intelligibility remained significantly different between -9 and -12 dB TMR (see Figure 6.5b).

We suggest that this result may be due to listener disengagement under the most adverse conditions in English. Previous research has shown that disengagement can occur when task difficulty is thought to be beyond a perceived level of achievability (Hopstaken et al., 2015). This result also illustrates the fall in intelligibility in a non-native language compared to a native one. We suggest that the same effect could have been observed in French if we had tested more adverse TMR levels (see Figure 6.5b).

4.3 The effect of language on energetic and informational masking

In SIN and SIS conditions, masking affects intelligibility at two distinct, but simultaneous processing levels. In SIN conditions, pure energetic interference disturbs speech processing at the peripheral level. In SIS conditions, complementary semantic interference makes segregation and comprehension processes more difficult.

Nevertheless, our results demonstrate that intelligibility under adverse conditions becomes more challenging when speech is presented in a non-native language compared to a native language, whatever the type of interference. Moreover, our findings suggest that processing speech stimuli presented in a non-native language under adverse conditions increases cognitive load. This cost is observable both in terms of SI degradation and subjective LE.

Segregation between the target and masker was shown to be easier for native French listeners when the speech was presented in French rather than in English. Naturally, native listeners are experts, both at semantic and linguistic levels, as they are more exposed to this language. They may also be more likely to associate French voices with familiar characteristics (Johnsrude et al., 2013), compared to English voices.

Lecumberri et al. (2010) noted that non-native listeners face the dual challenge of an imperfect signal and imperfect knowledge. Similarly, in SIN scenarios, non-native listeners have consistently shown greater difficulty with intelligibility than native listeners, due to energetic masking (Lecumberri and Cooke, 2006; Mattys et al., 2012; Rogers et al., 2006).

In the present study, our participants knew some English (mean LEAP-Q score 7.01/10) but were not native speakers. Other studies have used unfamiliar language to assess the influence of language on intelligibility in complex auditory environments (Brouwer et al., 2012; Lecumberri and Cooke, 2006; Van Engen and Bradlow, 2007). For example, Brouwer et al. (2012) showed that an unfamiliar language mask had a relatively small impact on intelligibility compared to a familiar or native language.

The intelligibility differences we observed between languages in adverse conditions are corroborated by our results for subjective LE. In both SIS and SIN conditions, participants reported that tasks that used the English version of the CRM corpus were more effortful than those that used the French corpus. The interaction between language, energetic and informational masking supports the idea that speech segregation requires more cognitive resources in a non-native language—even for simple words. In our study, intelligibility in adverse conditions (e.g., -9 dB SNR or TMR) was, on average, similar in SIN and SIS conditions in French. However, this was not the case for the English corpus. Here, the percentage of correct responses decreased to 20% in the SIS condition, while it stabilized at around 50% in the SIN condition. This difference may be due to additional informational processing in the SIS condition. As Van Hedger and Johnsrude (2022) noted in their review of speech perception in adverse conditions, "different adverse conditions place different demands on cognitive resources."

Based on the results presented here, we suggest that the "knowledge-driven" processes described by Mattys et al. (2012), which listeners rely on to perceive speech in complex auditory situations, are more fragile when operating in a non-native language, and come with a higher cognitive cost (Golestani et al., 2009; Lecumberri and Cooke, 2006). This difference is clearly illustrated by the better intelligibility of the French version of the CRM corpus compared to the English version in the most adverse conditions. Finally, we suggest that the difference described above is mainly due to the increased cognitive demands involved in comprehending and retrieving information when processing speech in a non-native language in complex conditions.

Furthermore, the observation that intelligibility differences between the two languages are inhomogeneous across masking levels suggests that using a foreign language to assess SIN and SIS processing may be less reliable than using participants' native language. If the observed differences between French and English had been significant and homogeneous across all masking levels, the impact would be more predictable—using an English rather than a native-language corpus would simply be associated with lower average performance and increased LE. However, the present results indicate that the impact of the use of an English corpus on intelligibility and LE is most pronounced under adverse conditions, particularly those involving high levels of masking. This results in heterogeneous costs across TMR and SNR levels. Consequently, we argue that the loss of intelligibility in a non-native language is related to increased LE, but this is highly dependent on the complexity of the auditory environment.

4.4 Language proficiency

The purpose of this study was to evaluate the impact of using an English-language compared to a French-language CRM corpus on SI and LE in a sample of native French listeners. English proficiency was subjectively assessed to ensure that participants understood the stimuli. Although we made no hypotheses about English proficiency levels, it is important to acknowledge that language proficiency significantly affects intelligibility in non-native listeners, especially under complex auditory conditions (Cooke and Lecumberri, 2012; Lecumberri and Cooke, 2006; Rhebergen et al., 2005; Rogers et al., 2006; Warzybok, Zokoll, et al., 2015).

For example, Smiljanić and Bradlow (2011) showed that language proficiency affects intelligibility, and noted that more experienced non-native listeners are better able to cope with background noise than less experienced ones. Similarly, Warzybok, Brand, et al. (2015)demonstrated that both language proficiency and the type of speech task have a strong influence on intelligibility. However, given the restricted set of response alternatives and the simplicity of the corpus language, the CRM task used in our study may be less sensitive to differences in proficiency.

4.5 Limits and Perspectives

This study could have been strengthened by including an objective assessment of English proficiency, based on more robust tests. Including language proficiency as an additional factor in our analyses would have provided a more nuanced interpretation of the results. For instance, standardized English tests such as the TOEFL (Educational Testing Service, 2024) or IELTS (British Council, IDP: IELTS Australia & Cambridge Assessment English, 2024) could have been used.

Further studies could extend this work by investigating other foreign language pairs, as this would help to assess the reliability of the findings presented here and explore whether effects may differ depending on the corpus language.

5 Conclusion

This study investigated the influence of native (French) and non-native (English) languages under conditions of energetic and informational masking. The main result is that differences in SI and LE between the two languages are not homogeneous across task difficulty but rather depend on the relative masking level.

These results are of interest for the design of the next studies of this project, as well as further studies. They could be used to investigate LE in complex auditory situations, particularly when incorporating objective electrophysiological measures, as we do in the next steps. The observed decline in intelligibility in English, compared to French, under the most adverse conditions, supports the hypothesis **H1** that using a non-native language may lead to information loss for non-native listeners, even with simple auditory stimuli.

Publication

This contribution led to an article currently in press at Ear and Hearing and a Poster (see Appendix **??**) presented at the "Journées Perception Sonore"(2023, Paris) Conference. Data are available on the following OSF repository: https://osf.io/ye4f7/overview

7

Executive Functions & Challenging Listening

Description of the Chapter

In a second study, the role of EFs and neural mechanisms in listening in challenging environments was explored. It first examined how high-level cognitive processes, such as updating and inhibition, interact with SI and LE across varying levels of environmental difficulty. We then investigated potential neural correlates of LE, with a focus on alpha oscillations measured by EEG, highlighting how difficulty levels in SIN and SIS conditions could modulate these dynamics. This chapter aims to provide a comprehensive view of the cognitive and neural mechanisms underlying listening in challenging auditory environments.

Hypothesis 2

Executive functioning (especially inhibition) would be correlated with speech intelligibility and with listening effort.

Hypothesis 3

EEG alpha oscillation dynamics would be impacted by the complexity of the auditory scene.

Main Results

In SIN, no relation was observed between SI and EF performances, nor between subjective LE and EF performance.

In SIS, participants with better scores in the updating dual N-back task had better SI in all SIS conditions.

In SIS, participants with better performances in the inhibition Stroop task had better SI in the most adverse SIS condition.

In the time-frequency domain, alpha power increased in the left temporo-parietal region in both SIN and SIS conditions.

This alpha power change was associated with different independent components.

These components showed either desynchronization or synchronization, confirming the existence of different alpha generators in challenging listening.

1 Research goal

Understanding speech in a complex auditory environment involves both auditory and non-auditory mechanisms (Eckert et al., 2016). High-level cognitive functions such as EFs are also involved in this process, and their relationship with SI and LE is valuable to understand the general mechanisms underlying listening (Francis and Love, 2020). Thus, in this study, the core EFs described in the literature by Miyake et al. (2000), including updating, shifting, and inhibition, were investigated in relation to LE and SI.

LE is a multidimensional concept that, beyond auditory perception, relies on higher-order cognitive mechanisms. When an auditory scene becomes complex, listeners implement strategies to extract the relevant information from the auditory streams. They analyze the auditory scene and segregate the different auditory streams using various cues such as, for example, differences in sound level between talkers, spatial localization, or gender of the talkers. Once the streams are segregated, listeners must select the stream of interest while ignoring the others. Here, inhibition is hypothesized to play a key role. Strong inhibition abilities have been associated with better SI, while lower inhibition has been associated with increased LE (Perrone-Bertolotti et al., 2017; Stenbäck et al., 2016). By inhibiting the masker talkers or the noise, the listener can better select the target stream and access to semantic content of the speech. Furthermore, the variety of strategies that listeners can implement during listening in complex environments suggests that different cognitive processes are involved. In particular, in multi-talker situations, using differences in target and masker sound levels as a cue for stream segregation may rely on inhibitory control (Lanzilotti et al., 2022; Stenbäck et al., 2016). Moreover, working memory, which is closely related to updating, has been associated with LE (Rudner et al., 2011) and SI (Besser et al., 2013; Ingvalson et al., 2017) in complex auditory situations, suggesting that better capabilities to maintain auditory elements in working memory may help improving SI and reducing LE in challenging auditory environments.

In this study, our objective was to assess these relationships by comparing SI and subjective LE with EF performances. In particular, we investigated whether this relation varies depending on the difficulty level of SIN and SIS conditions. We selected nine EF tasks commonly used in the literature Chenot et al., 2024; Miyake et al., 2000 to measure the EF performances regarding Miyake et al. (2000)'s model.

This study aims to address the second hypothesis of this project:

> **Hypothesis 2**
>
> Executive functioning (especially inhibition) would be correlated with speech intelligibility and with listening effort.

This hypothesis can be investigated through the following complementary hypotheses:

Complementary Hypotheses - Behavior

H2.1a : Reproduction of study 1 and literature: SI and LE would be impacted by the masking level in both SIN and SIS.

H2.1b : In comparison to SIN, the presence of IM would reduce SI and increase LE in SIS.

H2.2a : Performance in executive functions, especially inhibition, would be correlated with SI in SIN and SIS scenarios.

H2.2b : Performance in executive functions, especially inhibition, would be correlated with LE in SIN and SIS scenarios.

In addition to this first goal, we also explored complementary neural correlates of listening in effortful auditory situations, focusing specifically on alpha dynamics. The search for an objective way to measure LE remains a central question in speech research. When assessing LE with subjective questionnaires, self-reported results can be influenced by factors such as participants' motivation, experience with the task, or perception of their own performance. Consequently, subjective LE is difficult to interpret and can vary widely between participants. This is why objective measures of LE could provide new insight into the underlying mechanisms. Many methods are used to investigate objective measures of LE, including pupillometry, fNIRS, fMRI, MEG, or EEG.

Within the scope of this project, EEG was chosen as the tool to investigate potential neural correlates of LE. Alpha oscillations have been described in relation to listening in complex environments in *a priori* contradictory ways. Some studies report increases in alpha power during effortful listening (Obleser et al., 2012; Wilsch et al., 2015), while others describe a decrease in alpha power (Dimitrijevic et al., 2017). One explanation for these inconsistencies seems to rely on the presence of multiple alpha generators in the brain (Wisniewski, 2017). In this study, we investigated alpha oscillations across different difficulty levels in both SIN and SIS conditions. The goal of this investigation was to address the third hypothesis of this project:

Hypothesis 3

EEG alpha oscillation dynamics would be impacted by the complexity of the auditory scene.

Spectral and time-frequency analyses were conducted to assess cortical responses across different levels of auditory difficulty. As reported in the literature Wisniewski and Zakrzewski, 2023; Wisniewski et al., 2024, interpreting alpha oscillations in the time-frequency domain is challenging. Therefore, analyses were extended by extracting independent components of the EEG signal to determine whether the observed alpha dynamics corresponded to different components of the EEG signal. To further characterize these components, exploratory source localizations were performed.

2 Material and Methods

2.1 Participants

30 native French participants were recruited for this experiment (15 women, 15 men, mean age 25.9 $\pm$ 3.4 years old). Their audition was controlled by pure-tone audiometry using an Elios® clinical audiometer (Echodia, Le Mazet-Saint-Voy, France), for the following frequencies: .25, .5, 1, 2, 4, 6, 8, and 12.5 kHz (hearing level $\leq$ 20 dB; mean hearing level .9 $\pm$ 8.3 dB HL, details in Appendix 10.1). To be included in the experiment, participants had to meet the following criteria: aged between 18 and 40 years, native French speakers, no known vision or hearing impairments, no medication targeting the central nervous system, no neurological or psychiatric conditions, and no known brain lesions. Participants could not be included if they had participated in the previous studies. They received €25 as financial compensation for their participation. This study was approved by the local ethics committee of Toulouse (IRB Number 2024-817). All participants provided written informed consent prior to data collection.

Sample Size calculation

28 participants from the previous study (see Chapter 6) had also participated in another study conducted in the laboratory (Chenot et al., 2024), which (among other tasks) included the same EF evaluation tasks. Their data were merged to compare SI and EF performances for the estimation of the sample size required for this second study. A significant Pearson correlation was found between the global EF Z-score and mean performance in adverse SIS conditions (below 0 dB TMR), with $r = .52$ and $p = .0065$ (see Figure 10.1 in Appendix).

Based on this correlation, a power analysis was conducted using the *pwr R* package, with parameters set to $r = .05, \alpha = .05$, and $power = .8$. The results indicated that a minimum of 28 participants would be required to achieve sufficient statistical power for the study. The participants used for the sample size calculation were not included in the present study.

2.2 Stimuli

Speech perception

As suggested by the first study (Chapter 6), the Coordinate Response Measure (CRM) corpus was used in its French version (Isnard et al., 2024). The advantage of using a corpus in the native language of the participants was previously described using this exact version of the CRM. The original corpus used in the first experiment included digits from 1 to 8. The present study restricted the corpus to sentences including digits from 1 to 4, to balance with the four available colors, ensuring equal statistical probabilities for colors and digits. The 128 resulting combinations (8 call signs x 4 colors x 4 digits) led to a total of 1024 stimuli (8 call signs x 4 colors x 4 digits x 8 speakers).

Similarly to the previous experiment, the CRM was used in SIS and SIN conditions and a control condition with a single target talker was presented without any masking signal (i.e., target alone), to confirm that participants were able to understand speech clearly in the absence of interference.

Executive Functions

EFs were assessed using 9 different tasks, as described below. The tasks were displayed on a screen in front of the participant. The Inquisit software[1] was used to present the EF tasks.

Updating:

Letter Memory (Friedman et al., 2008): A series of 5, 7, or 9 letters appears on the screen one at a time, which the participant must progressively memorize while saying them out loud. At the end of the series, the participant is required to click on the last three letters they memorized.

Dual N-back (Jaeggi et al., 2010): A blue square appears on the screen in one of 8 possible locations, simultaneously with one of 8 possible letters. For N = 0, the participant must determine whether both the location of the square matches the target (press “A”) and the displayed letter matches the target letter (press “L”). For N = 1, the participant must determine if the location and the letter match those from the previous trial. For N = 2, the same instructions apply as for N = 1, but the participant compares the stimuli to those presented two trials earlier.

Keep Track (Friedman et al., 2008): Words belonging to one of six possible categories (e.g., animal, color, etc.) appear on the screen one at a time. The participant is asked to track 2 to 4 target categories among the displayed words and progressively remember the last word from each category, typing it out on the keyboard.

Shifting/Switching:

Category Switch (Friedman et al., 2008): Words are presented on the screen. The participant must determine whether the word refers to a living or non-living object, or whether it is smaller or larger than a basketball by clicking on the corresponding category with the mouse. From one trial to the next, the task can either be the same or different. A ”switch” refers to trials where the dimension changes from one to the other.

Color-Shape (Miyake et al., 2004): A shape (triangle or circle) appears on the screen on a red or green background. The participant must identify the color or shape of the stimulus by pressing the A and P keys on the keyboard. Either the instruction remains the same for an entire block, or it changes randomly from one trial to the next.

Number-Letter (Miyake et al., 2004): A 2x2 matrix is displayed on the screen. A pair of characters consisting of a number and a letter appears in each cell, one after the other, in a clockwise direction. When the stimulus appears at

[1] www.millisecond.com

the top, the participant must determine whether the letter is a consonant or a vowel. When the stimulus appears at the bottom, the participant must determine whether the number is even or odd by pressing the A and P keys on the keyboard. As the presentation of characters is predictable, the trials are of the switch type for all odd-numbered trials.

Inhibition:

Antisaccade (Friedman et al., 2008): The participant fixates on a cross at the center of the screen. A quick flash appears to the right or left of the cross, followed by a target stimulus on the opposite side. The target is an arrow pointing to the right, left, or up. The participant is instructed to press the key corresponding to the target that appeared on the screen.

Stop Signal (Verbruggen et al., 2004) A fixation cross is displayed at the center of the screen. When an arrow appears to the right or left of the cross, the participant must press the corresponding key. However, if a sound signal appears, the participant is instructed not to press any key.

Stroop (Stroop, 1935): Color names, written in color, appear one by one on the screen. The participant must say aloud the color in which the word is written as quickly as possible. The participants' voices will be recorded using a microphone. Congruent trials present a color name written in the same color as the word. Incongruent trials present a color name written in a different color. For control trials, a colored rectangle appears.

2.3 Apparatus

Audio and voice recording

Auditory stimuli were presented through ER-3C earplugs (Etymotic Research; see Figure 7.1), designed to minimize electromagnetic fields around the head, to avoid EEG interference, at a sound level of 55 dB SPL, ensuring clear perception without participant's discomfort. In addition, this system provides better sound isolation from possible surrounding noise that may occur in the lab's vicinity.

Electrophysiology

EEG data were recorded using a 64-active electrode system with a ®Biosemi Active II amplifier. Electrodes were installed using conductive gel (Signa gel) and placed according to the 10-20 international system.

2.4 Procedure

After providing informed consent, participants completed pure-tone audiometry. Then, the EEG cap was installed on the participant's head.

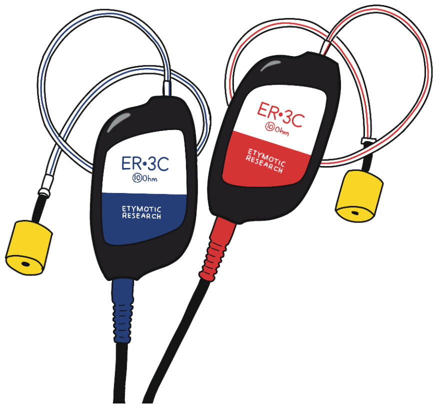


Figure 7.1: ER-3C headphones.

Resting State

After the EEG installation, participants completed a 5-minute resting state (RS), during which they alternated between 30 seconds with eyes open and 30 seconds with eyes closed, following on-screen and auditory instructions. During RS, participants were instructed to let their thoughts flow freely while avoiding falling asleep. At the end of the RS, they answered the Amsterdam Resting-State Questionnaire (ArsQ) to report their feelings and experiences during the RS.

Speech Intelligibility

Participants then completed the speech intelligibility task, using the CRM corpus (Section 1.1 in Chapter 5). In each experimental block, they were asked to indicate the color and digit they heard associated with the call sign "*Delta*" (in French sentences), by clicking the corresponding button on the virtual response box (see Figure 7.2).

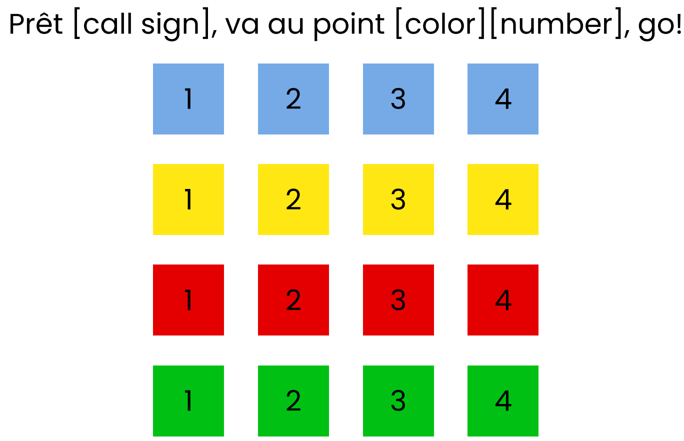


Figure 7.2: Coordinate Response Measure matrix answer presented on the screen.

The control condition with a single talker (one block of 16 trials) was always presented first to ensure the participants were able to do the task, followed by the SIN or SIS conditions, in a counterbalanced order. For both SIN and SIS conditions, 3 SNR or TMR levels were tested in three different blocks of 16 trials each (-9 dB SNR/TMR, -3 dB SNR/TMR and 3 dB SNR/TMR). In each trial, the talker(s) were randomly chosen to minimize talker-specific effects.

After each block, participants were asked to rate their subjective listening effort with the ESCU scale, rated from 0 (no effort) to 14 (only noise) (Section 1.2 in Chapter 5).

Each participant completed a total of 7 blocks, each containing 16 trials: 1 control block (16 unmasked trials) + 3 SIS blocks (16 trials x 3 TMR levels) + 3 SIN blocks (16 trials x 3 SNR levels). Each experimental block lasted approximately 1 to 2 minutes, and participants were allowed to take short breaks between blocks. The CRM tasks lasted for approximately 12 minutes in total.

Executive Functions

Following the speech intelligibility task, participants completed the 9 executive function tasks described in Section 2.2 in the following fixed order: Antisaccade, Letter Memory, Color-Shape, Number-Letter, Stroop, Keep Track, Dual N-back, Category Switch, and Stop Signal. In total, the EF tasks lasted approximately 1.5 hours.

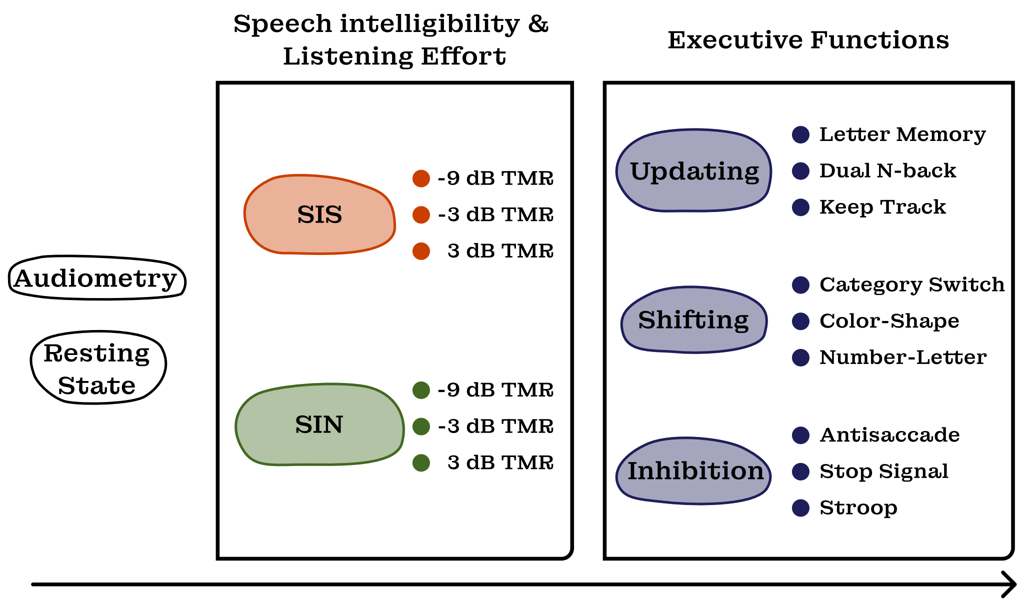


Figure 7.3: Schematic summary of the experimental procedure.

After EF tasks, the EEG cap was removed, and participants were given feedback on the study. The study lasted 2 hours in total.

2.5 Data analyses

The behavioral and electrophysiological data were analyzed using R 4.3.0 (statistics and behavioral plots), Matlab 2021a (EEG analyses), and Python 3.11.7 (speech-to-text Stroop extraction) and Loreta V20240713 (EEG source localization).

Behavioral Analyses

SI and LE

EF The Inquisit software provided a *.iqdat* file for each EF task of each participant, from where the EF metrics were extracted. The dependent variables of each EF are presented in Table 7.1.

EF	Task	Dependent variable	Metric
	Letter Memory	CR of recalled letters	
Updating	Dual N-back	CR	
	Keep Track	CR of recalled words	
	Category Switch	Δ RT	Switching cost
Switching	Color-Shape	RT	
	Number-Letter	Δ RT	Switching cost
	Antisaccade	RT	
Inhibition	Stop Signal	RT	
	Stroop	Δ RT	Inhibitory cost

Table 7.1: Dependent variables of each EF task. RT: Reaction time. CR: Proportion of correct response. Δ RT: difference of reaction time between congruent and incongruent trials (Inhibitory cost) or between switching and non-switching trials (Switching cost).

.

The performances of EF tasks were transformed to conform to normal distributions using Z-scores transformation with the *scale()* function in R and arcsine transformation on accuracy metrics (Keep Track, Letter Memory, Dual N-back). Reaction time Z-scores (Stroop, Antisaccade, Color Shape, Stop signal, and Number Letter) were inverted so that higher Z-scores corresponded to better performance. For example, in the Stroop task, lower Z-scores corresponded to higher inhibitory cost, indicating worse performance.

Stroop The Stroop task vocal responses were recorded with a RODE (NT-USB Mini) microphone and saved as .wav files. The Inquisit software does not allow for automatic detection and transcription of the voice in French. To address this, we used the whisper speech-to-text model (OpenAI) implemented in Python, as described in Section 1.4 in the Chapter 5.

CRM The percentage of correct responses was calculated as the number of trials in which both the color and digit were correctly identified, divided by the total number of trials in that experimental condition (16 trials in each block). Following current recommendations for speech analyses, scores were normalized using the Rationalized Arcsine Unit (RAU) transform in each condition (Studebaker, 1985).

Statistics For both SI and LE variables, a two-way ANOVA was carried out with within-subject factors Mask type (SIN *vs.* SIS) and Masker level (-9 dB *vs.* -3 dB *vs.* 3 dB SNR/TMR).

For SIN and SIS conditions, a one-way analysis of variance (ANOVA) with Greenhouse–Geisser correction for sphericity was conducted to examine the effect of masking level on SI and LE. The significance threshold was set at $p < .05$. Tukey's HSD (Honestly Significant Difference) post-hoc tests were applied for pairwise comparisons when appropriate.

ANOVAs were computed using the *aov_ez()* R function from the *afex* package. Pearson correlations using the *cor_test()* and FDR corrections were applied to compare EF performances with SI and LE. The significance threshold for all analyses was set at $p < .05$.

Electroencephalography Analyses

Preprocessing EEG data for spectral, time-frequency, and component analysis were preprocessed and processed as follows.

EEG data recorded during the CRM were extracted from the raw EEG files. Then they were average re-referenced, filtered on 4-25 Hz, and resampled at 512 Hz. Bad channels were removed and interpolated. Then, EEG data were split into data sets for each SIN and SIS condition, resulting in 9 data sets per participant. After that, mAMICA was applied to each condition data set, followed by independent component labeling with *ICLabel*. Eyes, muscles, and artifact components were flagged in order to be excluded from the signal. Data were epoched on a window of -1000 ms to 2000 ms around the CRM sentences onset. More details on EEG pre-processing methods in Section 3.1 of Chapter 5.

Time Frequency For the time frequency analysis (see Section 2.4 in Chapter), one-way ANOVAs with permutations and FDR correction were applied in SIN (SNR) and SIS (TMR) conditions using the EEGlab study statistics.

Independent Component analysis Component analysis allows additional exploratory observations of the EEG signal of the brain under SIS and SIN conditions. In this project, we aimed to reproduce the pipelines proposed in other studies (Jenson et al., 2015; Wisniewski et al., 2017, 2024). The component analysis (see Section 2.4) was performed on the component extracted with the mAmica algorithm, using the EEGlab study tool. For SIS and SIN separately, the component measures were precomputed, then a k-means clustering (with k = 13 and outlier clusters for IC clusters distant from more than 3 SDs from the centroid) was applied using time-based information (spectra and ERSPs) and location-based information (scalp maps). This clustering resulted in 13 components, which were analyzed per TMR or SNR condition, using one-way ANOVAs with permutation and FDR correction.

Source localization From the mAmica results, the source localization (see Section 2.4) of components of interest, revealed with the independent component analysis, were extracted using the LORETA software (v20240713). One-tailed sample t-tests (zero mean test A=0) were carried out on log-transformed sLORETA files.

Microstates Microstates (see Section 2.4) metrics of each participant were computed using the tool described in Section 3.2 in Chapter 5. Resting state EEG recordings were extracted from the raw EEG files. Then a re-reference was applied on Cz, high-pass (1 Hz), notch (50 Hz) filters, and resampling at 512 Hz. Then the MS pipeline was applied on participant and group levels (details in general methods). For clustering, k=7 clusters were selected, following Tarailis et al., 2024 recommendations. For the microstate analysis, one participant was excluded due to poor data quality of the resting state recording. Microstates results are presented in the Appendix 2.3.

3 Behavioral Results

3.1 Speech Intelligibility

For the sake of clarity, SI results in plots are presented as the percentage of correct responses; however, all statistical analyses were conducted on RAU-transformed data.

> **H2.1a - Behavior**
>
> Reproduction of study 1 and literature: SI and LE would be impacted by the masking level in both SIN and SIS.

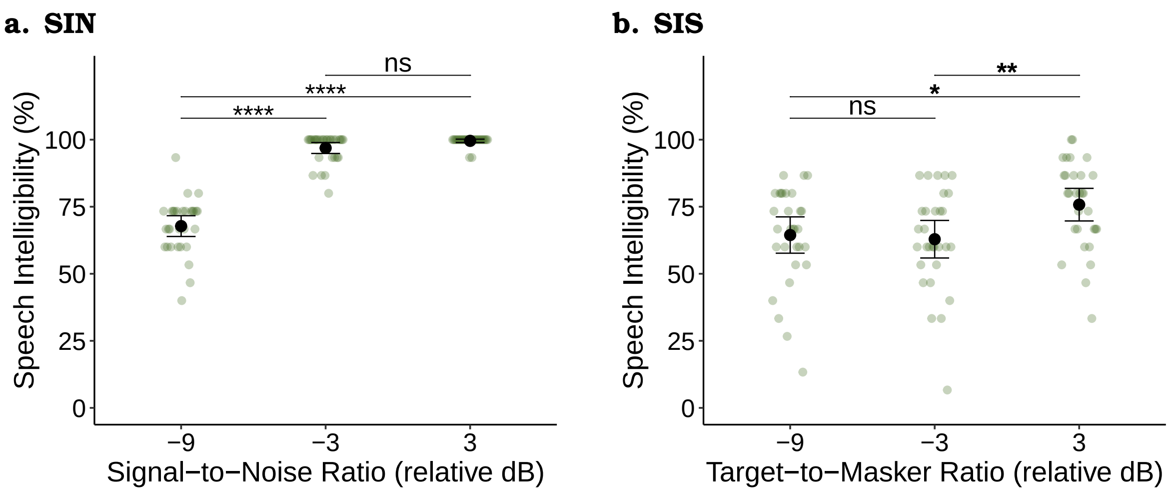


Figure 7.4: Mean speech intelligibility (in % of correct response) and confidence interval at 95% for each masking level of the SIN (a) and SIS (b) conditions.

Df: degrees of freedom, F: F-test value, η_p^2: partial eta squared
bold $p - values$ are significant

	Df	F	η_p^2	p
Level	2	106.6	.79	<**.001**
Mask	1	90.0	.76	<**.001**
Level:Mask	2	67.2	.70	<**.001**

Table 7.2: Two-way ANOVA for SI (RAU) in SIS and SIN conditions, with Level (-9 dB, -3 dB, and 3 dB) dB) and Mask (SNR, TMR) as factors.

The results of the two-way ANOVA, including Masking type (SIN *vs.* SIS) and Mask level (-9 dB *vs.* -3 dB *vs.* 3 dB SNR/TMR) are detailed in Figure 7.4 and Table 7.2. A significant interaction between Masking type and Mask level was found ($F(2, 50) = 67.2,\ p < .001,\ \eta^2 = .70$). Looking at the masking level effect within each mask type, the same analysis revealed higher SI complementary results. In SIN condition, SI was higher for 3 dB SNR than -9 dB SNR and higher for -3 dB SNR than -9 dB SNR. In SIS condition, SI was higher for 3 dB TMR than -3 dB TMR and for 3 dB TMR than -9 dB TMR.

H2.1b - Behavior

In comparison to SIN, the presence of IM would reduce SI and increase LE in SIS.

Post-hoc tests also showed that the Mask level effect was different between SIS and SIN with higher SI in SIN for -3 and 3 dB SNR/TMR levels, but not for -9 dB SNR/TMR.

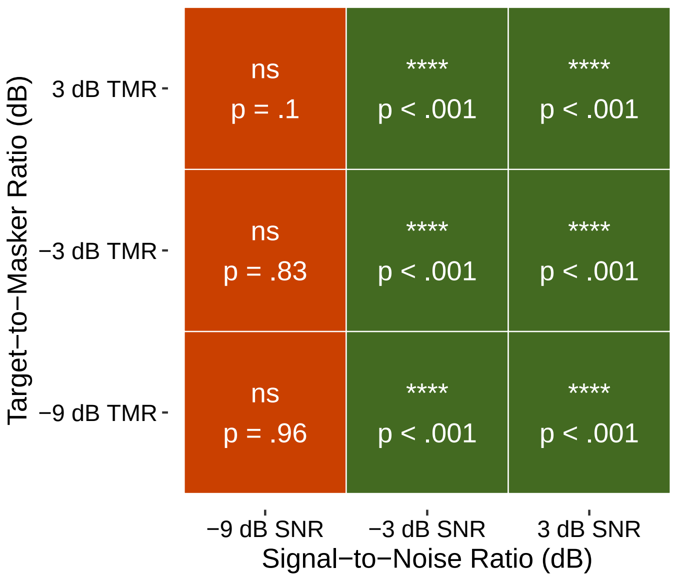


Figure 7.5: Post-hoc Tukey's HSD comparisons of SI (RAU) between SIS and SIN levels. Orange indicates non-significant differences; green indicates significant differences.

3.2 Subjective Listening Effort

H2.1a - Behavior

Reproduction of study 1 and literature: SI and LE would be impacted by the masking level in both SIN and SIS.

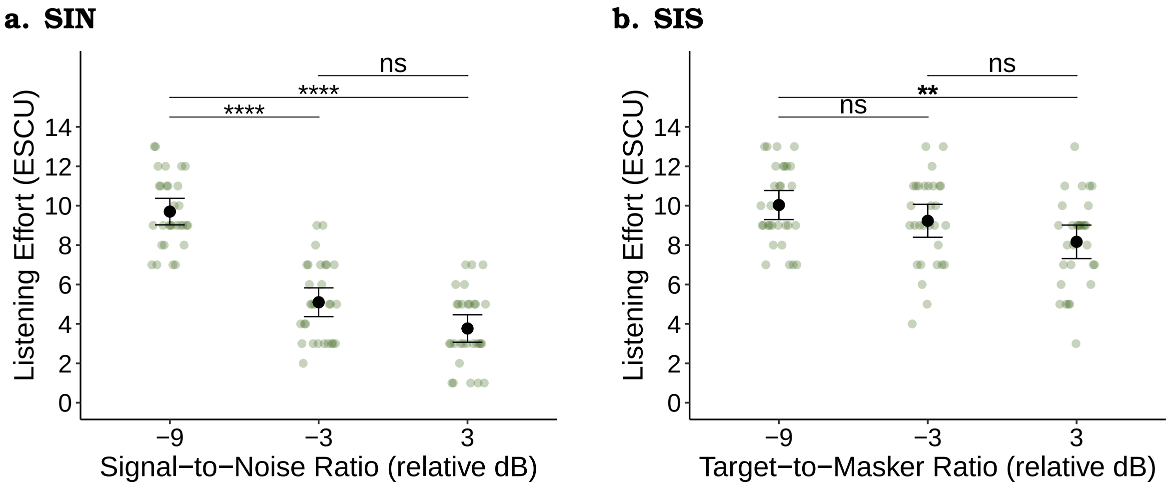


Figure 7.6: Mean subjective listening effort (ESCU) and confidence interval at 95% for each masking level of the SIN (a) and SIS (b) conditions

Df: degrees of freedom, F: F-test value, η_p^2: partial eta squared
bold $p - values$ are significant

	Df	F	η_p^2	p
Level	2	106.4	.79	<**.001**
Mask	1	102.2	.78	<**.001**
Level:Mask	2	62.5	.68	<**.001**

Table 7.3: Two-way ANOVA for subjective listening effort (ESCU) in SIS and SIN conditions, with Level (-9 dB, -3 dB, and 3 dB) dB) and Mask (SNR, TMR) as factors.

The results of the two-way ANOVA including Masking type (SIN *vs.* SIS) and Mask level (-9 dB *vs.* -3 dB *vs.* 3 dB SNR/TMR) are detailed in Figure 7.6 and Table 7.3. A significant interaction between Masking type and Mask level was found ($F(2, 50) = 62.5,\ p < .001,\ \eta^2 = .68$). Looking at the masking level effect within each mask type, the same analysis revealed higher LE complementary results. In SIN condition, LE was lower for 3 dB SNR than -9 dB SNR and lower for -3 dB SNR than -9 dB SNR. In SIS condition, LE was lower for 3 dB TMR than -9 dB TMR.

H2.1b - Behavior

In comparison to SIN, the presence of IM would reduce SI and increase LE in SIS.

Post-hoc tests also showed that the Mask level effect was different between SIS and SIN with lower LE in SIN for -3 and 3 dB SNR/TMR levels but not for -9 dB SNR/TMR.

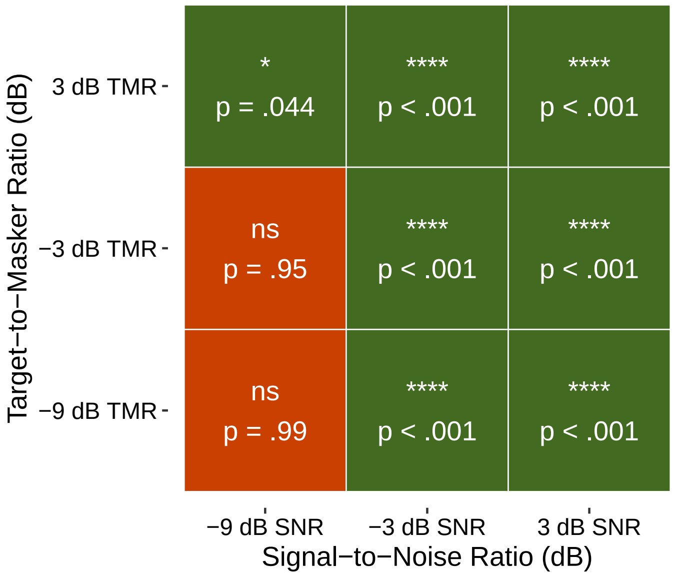


Figure 7.7: Post-hoc Tukey's HSD comparisons of LE (ESCU) between SIS and SIN levels. Orange indicates non-significant differences; green indicates significant differences.

3.3 Executive Functions and Speech Intelligibility

> **H2.2a - Behavior**
>
> Performance in executive functions, especially inhibition, would be correlated with SI in SIN and SIS scenarios.

Speech in Noise

The Pearson correlations between EF performances and SI for each SNR are presented in Figure 7.8, with the p-value and the r reported for each correlation. There were no significant correlations between any of the EF performances and the SNR (Figure 7.8).

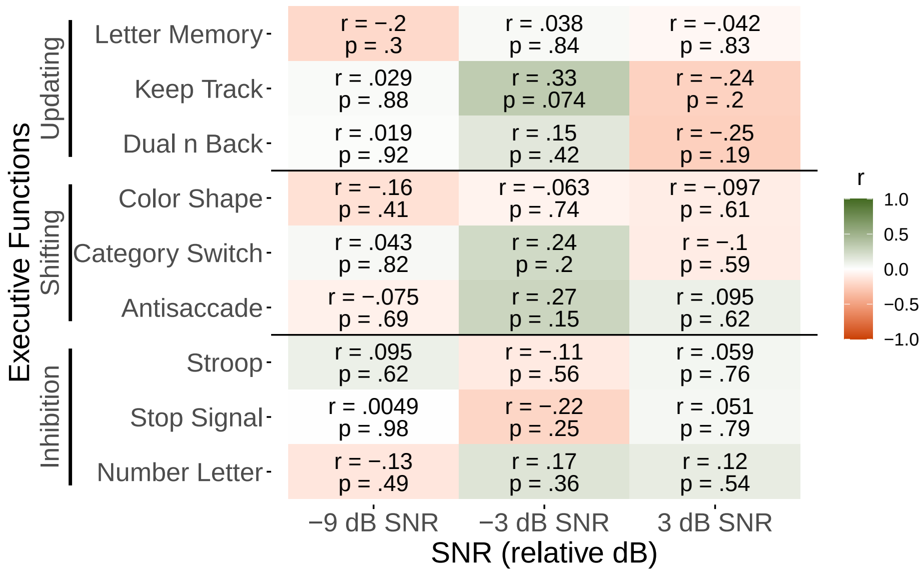


Figure 7.8: Pearson correlations between executive function task scores (Z-scores) and speech intelligibility in SIN. Significant correlations are shown in red and bold. p: p-value; r: correlation coefficient.

Speech in Speech

The Pearson correlations between EF performances and SI for each TMR are presented in Figure 7.9, with the p-value and the r reported for each correlation. This correlation indicates that SI increases as the Stroop Z-score increases, meaning the inhibitory cost decreases; consequently, the SI at -9 dB TMR also increases.

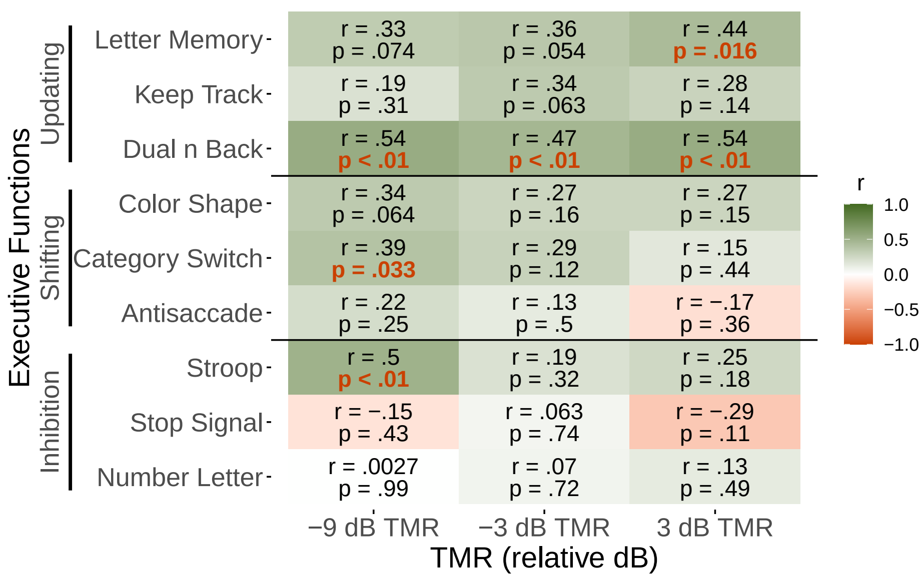


Figure 7.9: Pearson correlations between executive function task scores (Z-scores) and speech intelligibility in SIS. Significant correlations are shown in red and bold. p: p-value; r: correlation coefficient.

For updating, significant correlations were observed between the Dual N-back task score and SI at all three TMR levels (-9 dB TMR: $p < .01, r = .54$, -3 dB TMR: $p = .05, r = .47$, 3 dB TMR: $p < .01, r = .54$) and between the Letter Memory task score and SI at +3 dB TMR ($p = .02, r = .44$). A significant correlation between the shifting category switch cost and SI at -9 dB TMR was also observed ($p = .03, r = .39$).

For inhibitory tasks, a significant correlation was also observed between the Stroop score and SI at -9 dB TMR ($p < .01, r = .50$). This correlation is illustrated in Figure 7.10 in orange.

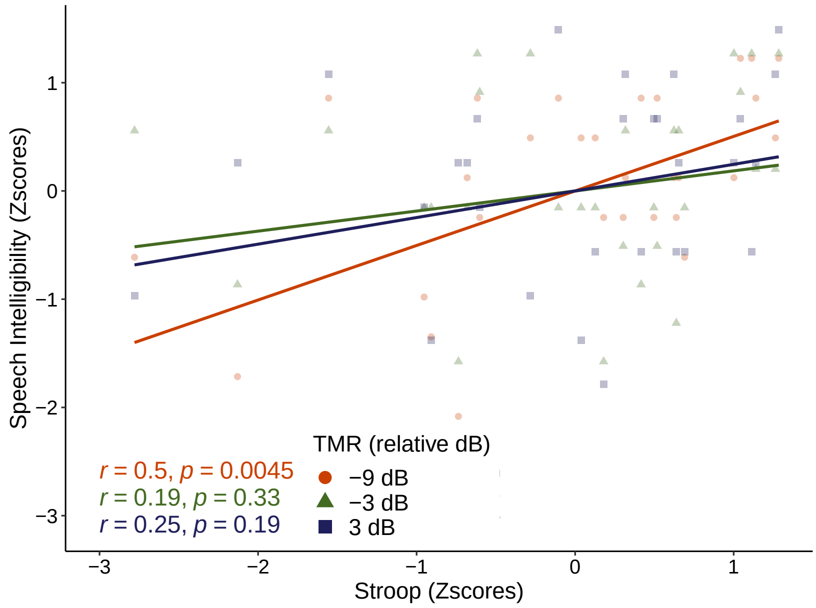


Figure 7.10: Pearson correlations between the Stroop inhibitory costs (Z-scores) and speech intelligibility in SIS (Z-scores RAU). p: p-value; r: correlation coefficient.

3.4 Executive Functions and Subjective Listening Effort

H2.2b - Behavior

Performance in executive functions, especially inhibition, would be correlated with LE in SIN and SIS scenarios.

Speech in Noise

The Pearson correlations between EF scores and subjective LE for each SNR are presented in Figure 7.11, with the *p-value* and the r reported for each correlation.

A significant negative correlation was observed between the updating Keep Track task and LE at -3 dB SNR ($p < .01, r = -.47$). No other significant correlations were observed.

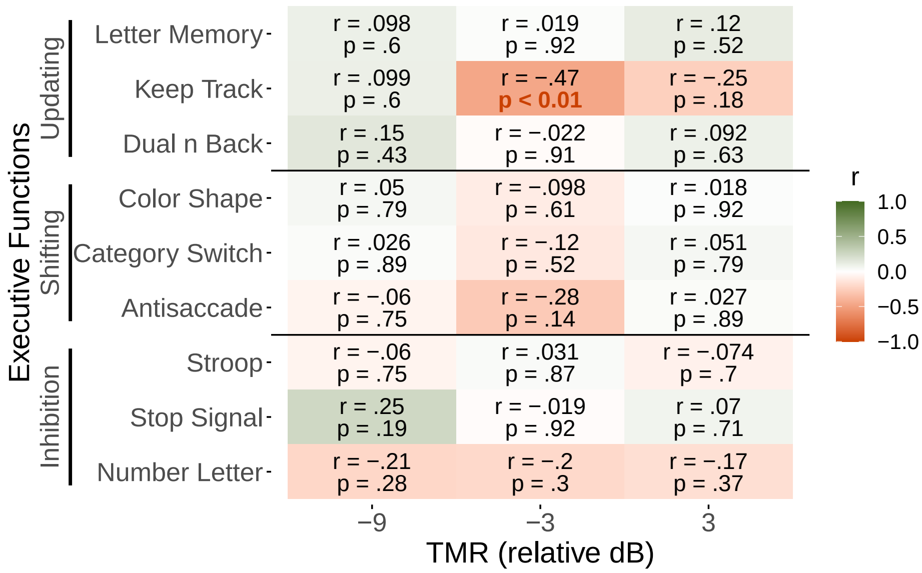


Figure 7.11: Pearson correlations between executive function task scores (Z-scores) and subjective listening effort (ESCU) intelligibility in SIN. Significant correlations are shown in red and bold. p: p-value; r: correlation coefficient.

Speech in Speech

The Pearson correlations between EF tasks and LE for each TMR are presented in Figure 7.12, with the p-value and the r reported for each correlation.

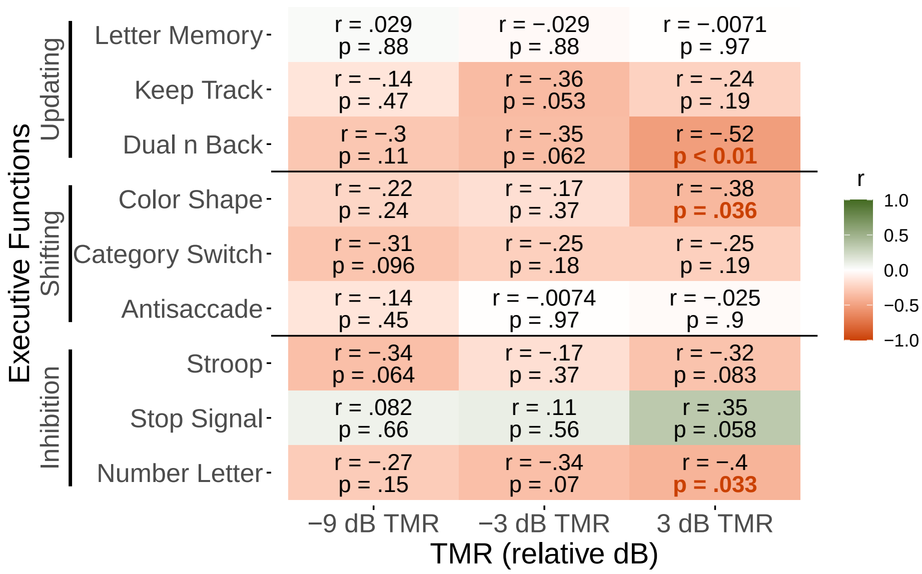


Figure 7.12: Pearson correlations between executive function task scores (Z-scores) and subjective listening effort in SIS. Significant correlations are shown in red and bold. p: p-value; r: correlation coefficient.

Significant negative correlations were observed between the updating Dual N-back task and LE at 3 dB TMR ($p < .01, r = -.52$). Additionally, a significant negative correlation was found between the Shifting Color Shape task and LE at 3 dB TMR ($p = .04, r = -.38$). No other significant correlations were observed.

4 Electroencephalography Results

4.1 EEG Spectrum

Alpha band

The mean alpha power (8-12 Hz) was computed for each SIN and SIS level. In SIN, significant differences in alpha power were identified around 12 Hz, with a higher alpha power at -9 dB SNR (see Figure 7.13a). In SIS, no significant differences in alpha power were observed between TMR levels (see Figure 7.13b).

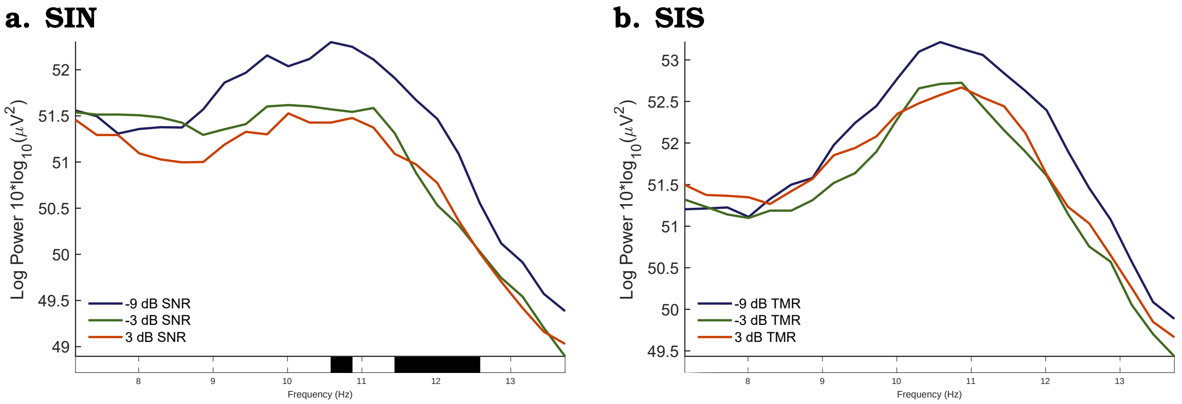


Figure 7.13: Mean alpha power across all participants for each SIN and SIS condition. Statistical significance assessed using FDR-corrected permutation tests ($p < .05$).

Theta

The mean theta power (4-8 Hz) was computed for each SIN and SIS level. In SIN, no significant differences in theta power were observed between SNR levels (see Figure 7.14a). In SIS, no significant differences in theta power were observed between TMR levels (see Figure 7.14b).

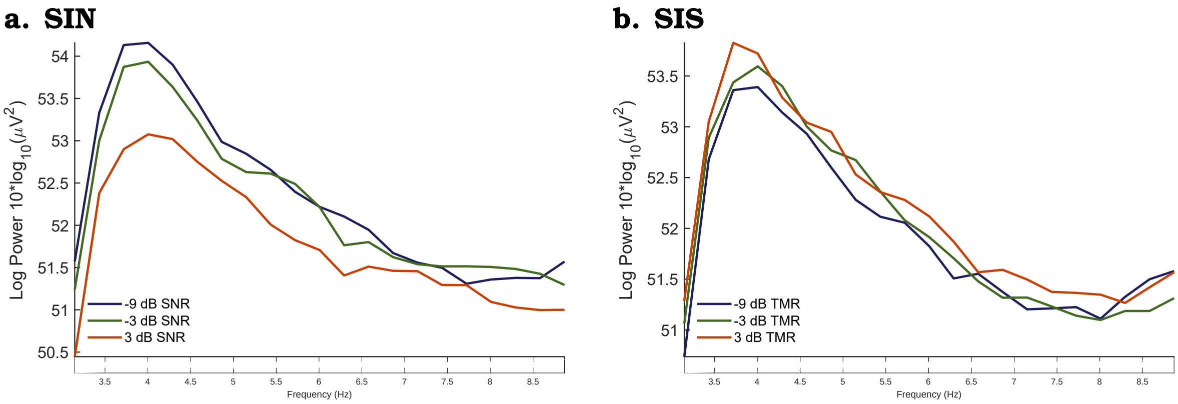


Figure 7.14: Mean theta power across all participants for each SIN and SIS condition. Statistical significance assessed using FDR-corrected permutation tests ($p < .05$).

4.2 Time Frequency

Speech in Noise

Alpha power topographic activity The Figure 7.15 illustrates the alpha activity (8-12 Hz) during the speech signal (0-2000 ms). Permutation test revealed a significant difference in alpha activity in the left temporo-parietal area between the different SNRs.

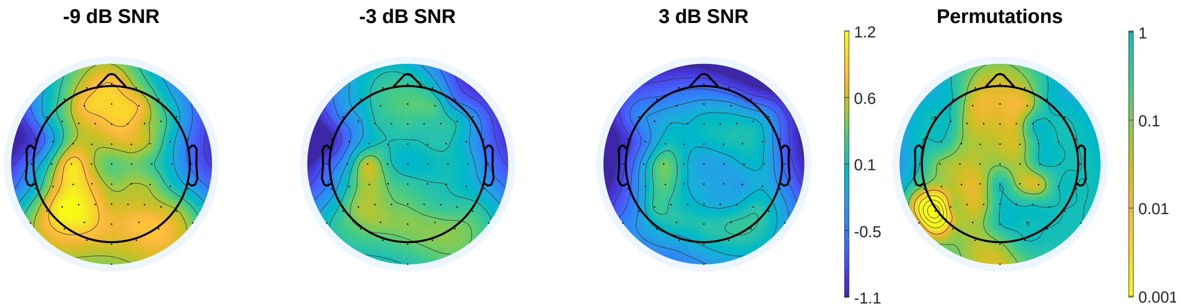


Figure 7.15: Topographical maps of mean alpha power (8–12 Hz; baseline-corrected) over the 0–2000 ms time window (0 = CRM sentence onset) in the SIN condition. The right panel shows the p-value map indicating significant differences between conditions.

Event-related spectral perturbation Based on previous topographic alpha activity showing a significant difference in the left temporo-parietal area, we selected T7, C5, C3, P7, P5, P3, TP7, CP5 and CP3 as a region of interest (ROI; see Figure 7.17), for subsequent analysis. Therefore, ERSPs were computed on this ROI for each SNR (see Figure 7.16). Permutation tests revealed significant differences between SNRs. More precisely, a significant difference between SNRs was found in the alpha activity and slightly above (8-13 Hz), starting around 1000 ms post-stimulus onset. This increase in alpha power appeared to be larger as the SNR decreased.

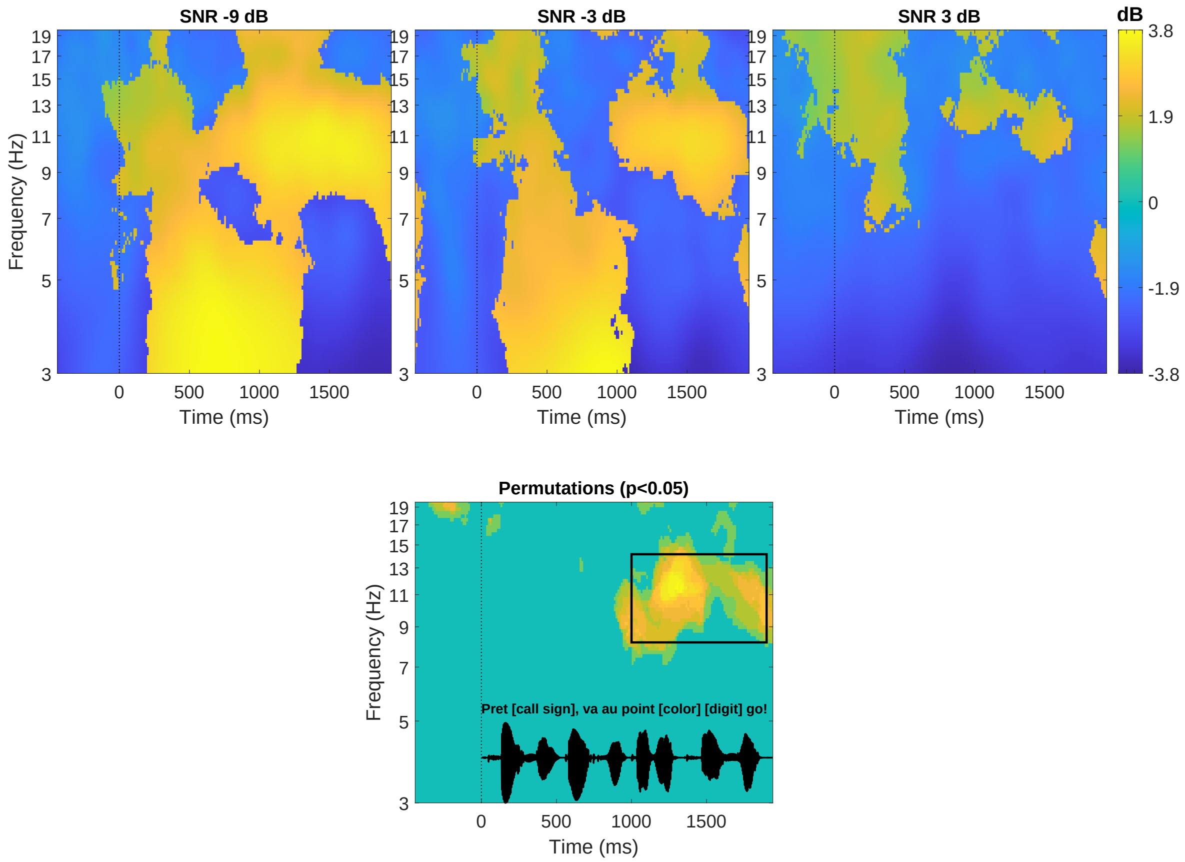


Figure 7.16: Time–frequency representations (ERSPs) for the left temporo-parietal ROI (T7, C5, C3, P7, P5, P3, TP7, CP5, CP3), showing spectral activity over time for each SNR level in the SIN condition, along with significance results (800 permutations; $p < .05$, FDR-corrected). Positive ERSP values indicate event-related synchronization (ERS), whereas negative ERSP values indicate event-related desynchronization (ERD). For illustration purposes, a generic CRM sentence spectrum is superimposed on the statistical time window (bottom).

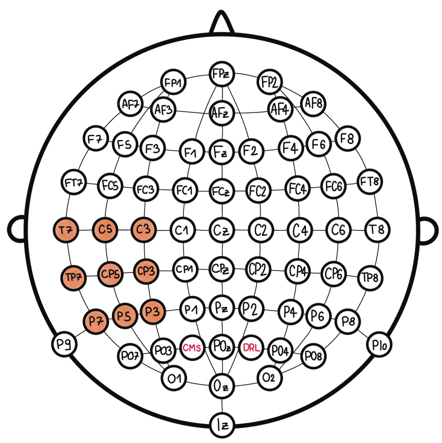


Figure 7.17: Left temporo-parietal region of interest (ROI) including electrodes T7, C5, C3, P7, P5, P3, TP7, CP5, and CP3.

Speech in Speech

Alpha power topographic activity The Figure 7.18 illustrates the alpha activity (8-12 Hz) during the speech signal (0-2000 ms). Permutations revealed a significant difference in alpha activity in the left temporo-parietal area between the different TMRs.

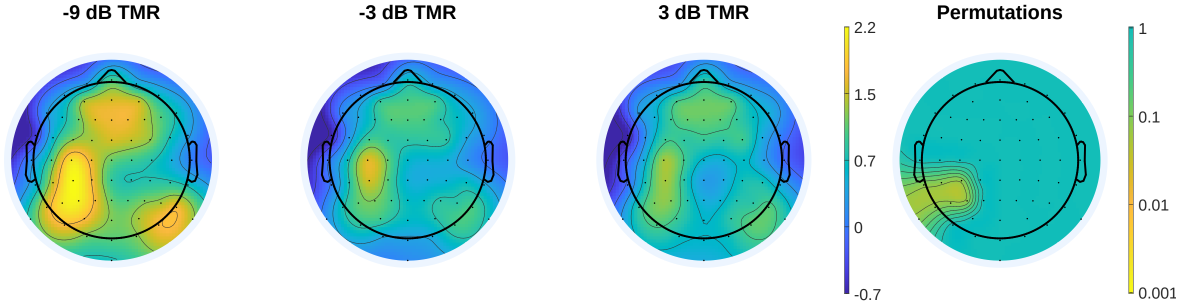


Figure 7.18: Topographical maps of mean alpha power (8–12 Hz; baseline-corrected) over the 0–2000 ms time window (0 = CRM sentence onset) in the SIS condition. The right panel shows the p-value map indicating significant differences between conditions.

Event-related spectral perturbation Like for the SIN analyses, we selected T7, C5, C3, P7, P5, P3, TP7, CP5 and CP3 as a region of interest (ROI; see Figure 7.17) corresponding to the left temporo-parietal region, for subsequent analysis. Therefore, ERSPs were computed on this ROI for each TMR (see Figure 7.19), and permutation tests revealed significant differences between TMRs. More precisely, a significant difference between TMRs was found in the alpha activity and slightly above (8-13 Hz), starting around 1000 ms post-stimulus onset with higher values for more adverse TMRs.

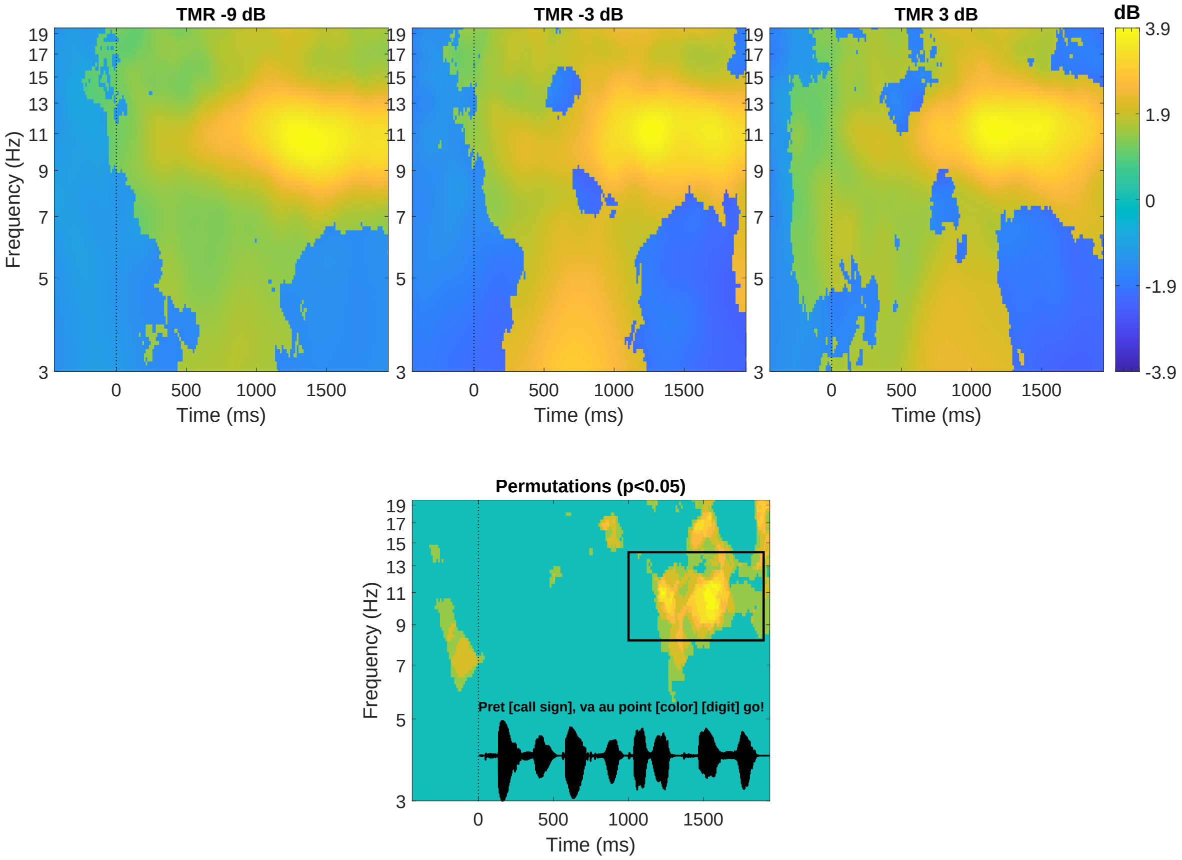


Figure 7.19: Time–frequency representations (ERSPs) for the left temporo-parietal ROI (T7, C5, C3, P7, P5, P3, TP7, CP5, CP3), showing spectral activity over time for each TMR level in the SIS condition, along with significance results (800 permutations; $p < .05$, FDR-corrected). Positive ERSP values indicate event-related synchronization (ERS), whereas negative ERSP values indicate event-related desynchronization (ERD). For illustration purposes, a generic CRM sentence spectrum is superimposed on the statistical time window (bottom).

4.3 Independent Component Analysis

Speech in Noise

Temporal Cluster In SIN, among the ICA clustering results, we found two components showing temporal distribution (see Figure 7.20). 141 components of 30 subjects were included in the left temporal IC cluster, and 185 components of 30 subjects were included in the right temporal IC cluster (see Table 7.4). In addition, the right and left temporal IC cluster associated ERSPs shows that a decrease in alpha activity during the speech processing was present. The right panel of Figure 7.20 also shows that a significant difference in alpha power around 11 Hz between SNRs for the left temporal IC cluster was present (that is, a higher value for -9 dB). For the right temporal IC cluster, no significant difference in alpha power was observed between SNRs. The sLoreta source localization showed a main source around the right Brodmann area 22 and 40 (Superior Temporal Gyrus and parietal lobe) for the left temporal IC cluster and around the left Brodmann area 2, 3 and 4 (Somatosensory and motor cortices) for the right temporal IC cluster.

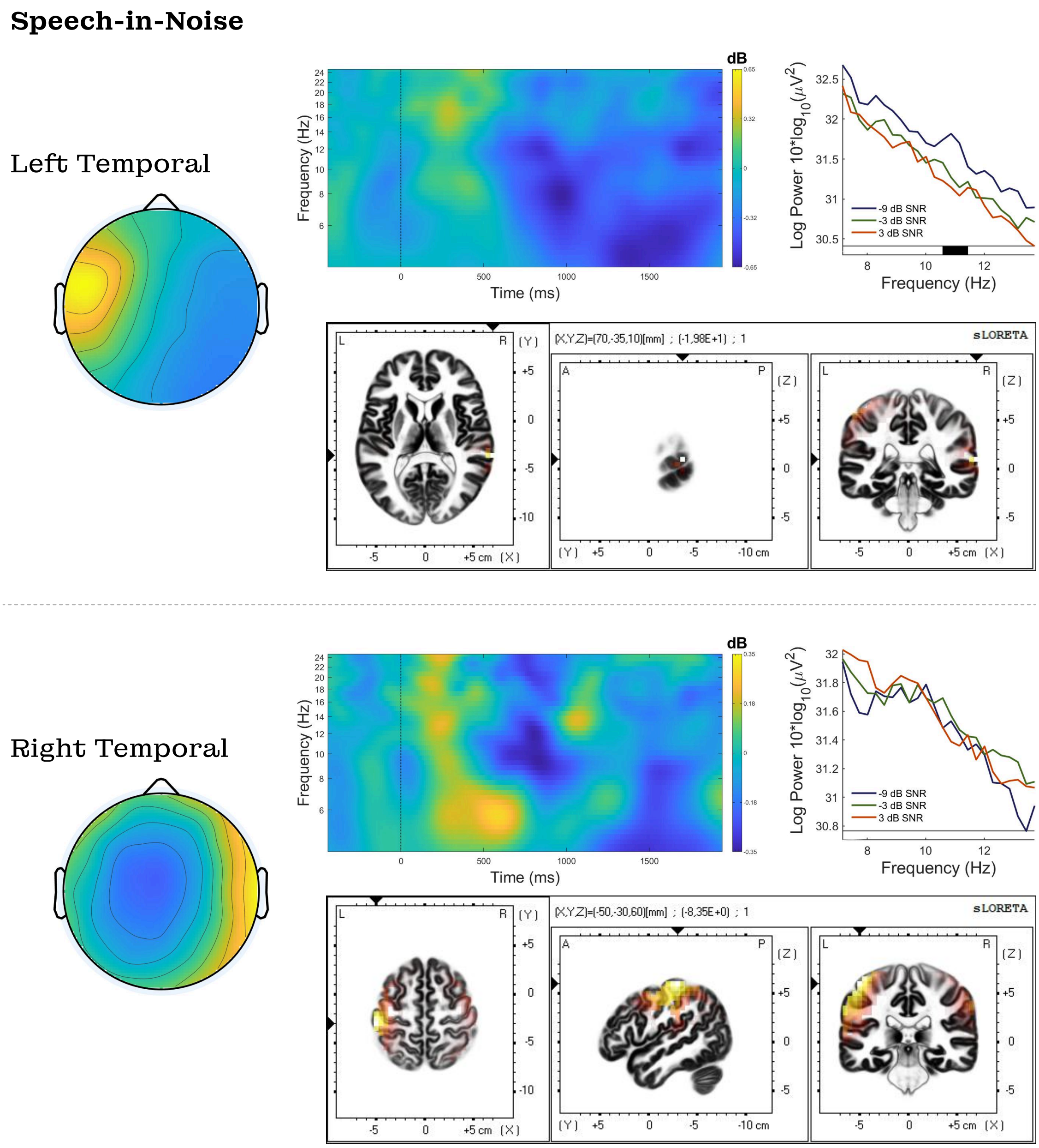


Figure 7.20: Average scalp maps of left and right temporal independent component clusters in the SIN condition. Left: average IC cluster scalp maps. Middle: ERSPs averaged across all SNR conditions. Right: alpha power for each SNR condition. Bottom: sLORETA source localization results. ERSPs for each SNR condition with corresponding statistical significance are provided in the Appendix 10.2).

Temporo-parietal Cluster In SIN, other components showed temporo-parietal (see Figure 7.21) activation. 182 components of 30 subjects were included in the left temporo-parietal IC cluster, and 140 components of 30 subjects were included in the right temporo-parietal IC cluster (see Table 7.4). The ERSP associated with the left temporo-parietal IC cluster showed an increase of alpha activity along the speech signal. The ERSP associated with right temporo-parietal IC cluster showed first a decrease, then an increase of alpha activity along the speech signal. No significant differences in alpha power were observed between SNRs for both temporo-parietal IC clusters. The sLoreta source localization shows a source in the right Brodmann area 42 and 22 (auditory cortex and Superior Temporal Gyrus) for the left temporo-parietal IC cluster and around the right Brodmann areas 2 and 3 (somatosensory cortex) for the right temporal IC cluster.

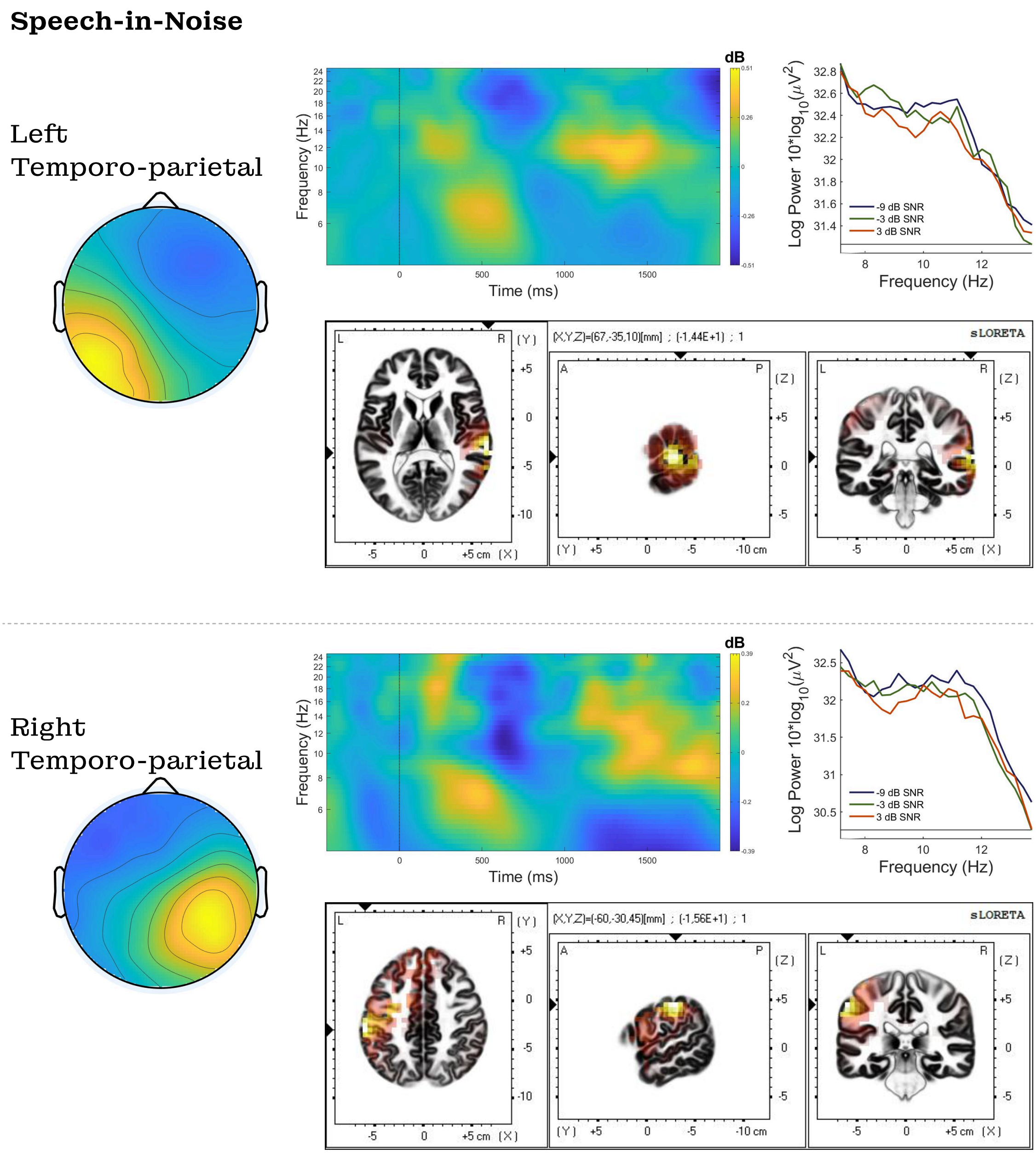


Figure 7.21: Average scalp maps of left and right temporo-parietal independent component clusters in the SIN condition. Left: average IC cluster scalp maps. Middle: ERSPs averaged across all SNR conditions. Right: alpha power for each SNR condition. Bottom: sLORETA source localization results. ERSPs for each SNR condition with corresponding statistical significance are provided in the Appendix 10.3).

Speech in Speech

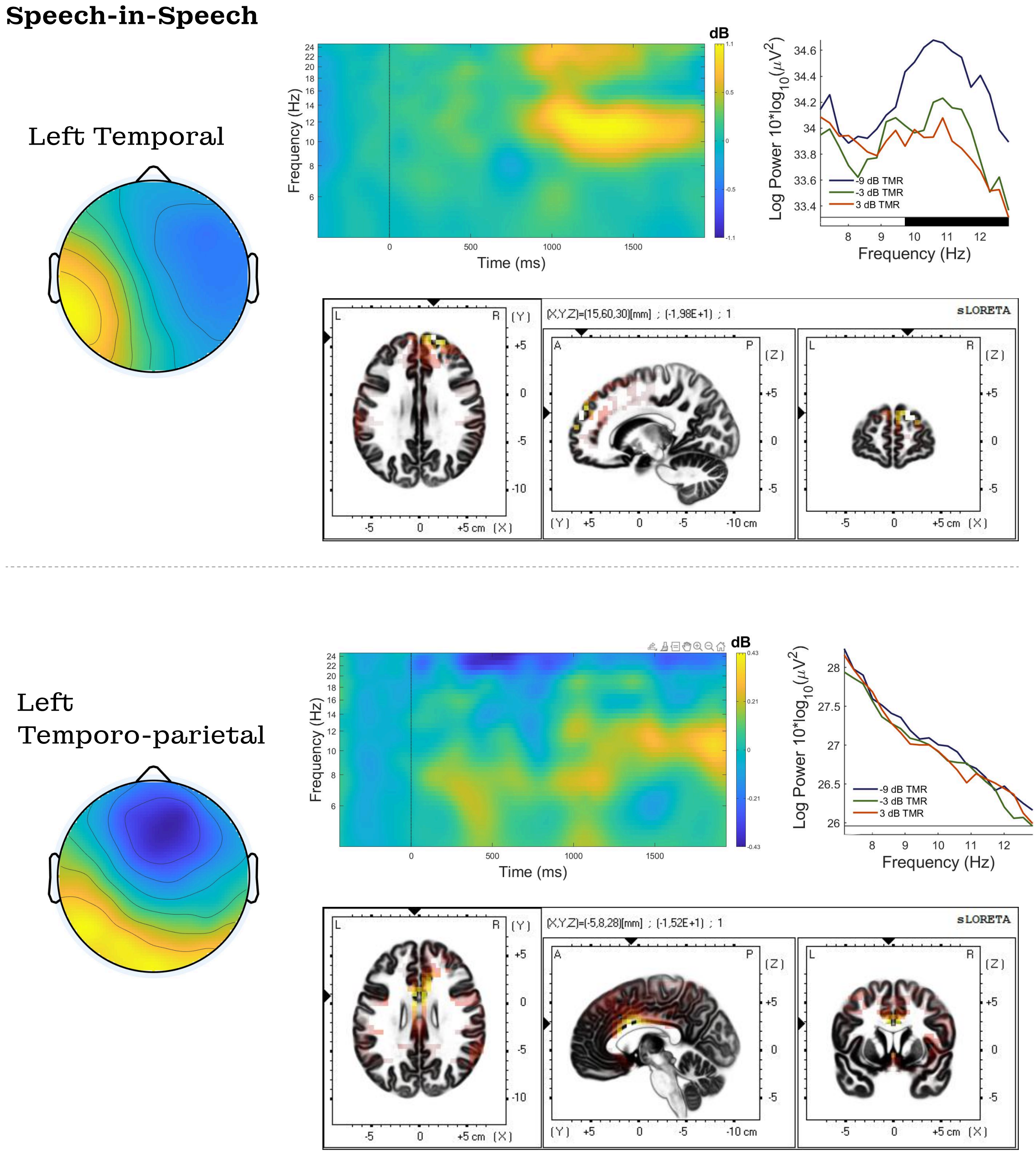


Figure 7.22: Average scalp maps of left temporal and temporo-parietal independent component clusters in the SIN condition. Left: average IC cluster scalp maps. Middle: ERSPs averaged across all TMR conditions. Right: alpha power for each TMR condition. Bottom: sLORETA source localization results. ERSPs for each TMR condition with corresponding statistical significance are provided in the Appendix 10.4).

In SIS, one IC cluster corresponded to left temporal activity and another to left temporo-parietal activity (see Figure 7.22). 140 components of 30 subjects were included in this left temporal IC cluster, and 158 components of 30 subjects were included in the right temporo-parietal IC cluster (see Table 7.4). The ERSP of the left temporal IC showed a low alpha activity during the speech signal, while the ERSP of the left temporo-parietal IC cluster showed increased alpha activity during speech processing. No significant differences were observed between TMRs. The sLoreta source localization showed a main source around the right Brodmann area 9 and 10 (Superior Frontal Gyrus) for the left temporal IC cluster and around the Brodmann area 24 and 33 (Limbic Lobe, Anterior Cingulate cortex) for the left temporo-parietal IC cluster.

Clusters composition

	Cluster	**Hemi**	**Subjects**	**ICs**	**Main sources**
SIN (n=30)	temporal	left	30	141	right BA 22/40
		right	30	185	left BA 2/3/4
	temporo-parietal	left	30	182	right BA 42/22
		right	30	140	left BA 2/3
SIS (n=30)	temporal	left	30	140	right BA 9/10
	temporo-parietal	left	30	158	BA 24/33

Table 7.4: IC Cluster compositions with number of subjects, number of included ICs and main cortical sources. BA: Brodmann area.

5 Discussion

The outcomes of this study provide several key insights. Here, we discuss the main results, first at the behavioral level, then at the electrophysiological level. A first goal of this study was to investigate the relationship between EFs and SI and LE, to address the following hypothesis:

> **Hypothesis 2**
>
> Executive functions performances (especially inhibition) would be correlated with speech intelligibility and with listening effort.

Additionally, this study aimed to answer the question of the possibility of finding LE markers with EEG. We addressed this with the following hypothesis:

> **Hypothesis 3**
>
> EEG alpha oscillation dynamics would be impacted by the complexity of the auditory scene.

5.1 The CRM corpus for speech intelligibility and listening effort assessments

> **H2.1a - Behavior**
>
> Reproduction of study 1 and literature: SI and LE would be impacted by the masking level in both SIN and SIS.

> **H2.1b - Behavior**
>
> In comparison to SIN, the presence of IM would reduce SI and increase LE in SIS.

First, we reproduced (Isnard et al., 2024) and our previous study (Chapter 6) results, validating the effectiveness of the CRM French version. Indeed, using this version, we replicated well-established findings with the CRM corpus in SIS and SIN conditions (Brungart and Simpson, 2007). As expected, participants demonstrated greater SI under favorable conditions compared to more adverse ones.

Looking more closely at the SIS condition, the difference in SI was significant, particularly between the less difficult condition (+3 dB TMR) and the two others, but not between the two more adverse conditions (-6 dB TMR and -3 dB TMR). Looking at other studies using the CRM in SIS with two simultaneous talkers (see Brungart, 2001a and according to our previous study results, see Chapter 6), the SI at -9 dB and -3 dB are in agreement with the current ones, showing a mean correct response stabilizing around 60%. Studies with even more TMR levels (Andéol et al., 2017; Brungart, 2001a; Lanzilotti et al., 2022, 6) show that this stabilization occurs at negative levels. This dynamic can be related to differences in listening strategies between participants (Andéol et al., 2017; Lanzilotti et al., 2022). In fact, in SIS adverse conditions, when the masker talker speaks louder than the target, some participants are able to use the sound level difference as a cue to segregate the streams, while others are not. This difference in strategy can explain how the mean performances of negative TMR do not show the same degradation as in SIS. Some participants perform in the same way in SIS and SIN, while others show an increase of SI for negative TMRs in SIS. Thus, in SIN, the presence of a single talker makes it impossible to rely on this type of strategy.

In parallel to SI, we collected subjective LE for each TMR and SNR. Similarly to SI and as expected, in both SIN and SIS scenarios, LE was related to task difficulty, increasing as the TMR or SNR became more defavorable (-9 dB SNR or TMR). Together with previous work (Krueger, Schulte, Zokoll, et al., 2017; Lanzilotti, 2021; Rennies et al., 2019), we can confirm that subjective LE in SIS and SIN is affected by the relative sound intensity between the target and the mask in a French version of the CRM. However, we observed that subjective LE varied differently in SIN and SIS scenarios. In SIN condition, LE was significantly different between all levels: the worse the SNR, the higher the subjective LE. In contrast, in SIS condition, a significant difference was observed only between the most extreme TMR (-9 dB and +3 dB TMR). We argue that this difference could be partly explained by distinct types of masking. In the SIS scenario, IM adds to EM, while in SIN, only EM is present. It is important to note that the order of tasks was counterbalanced through the experiment to ensure that the observed differences were not attributed to an order effect. As noted by Brungart (2001b), when using speech-shaped noise, masking is mainly energetic, while in SIS condition, the additional IM tends to have a dominant effect. This difference in masking affects not only SI, as discussed previously, but also the subjective LE, as supported by the data of this study.

Self-reported LE is, however, subjective (by definition) and therefore does not necessarily represent the same aspects of cognitive processing as objective LE measures (Zekveld et al., 2010). This distinction highlights the importance of investigating physiological correlates of both performance and effort, such as EEG neural markers. Furthermore, as introduced in Chapter 2.3 and stated by Winn and Teece (2021): "LE is not the same as SI". The ability of a participant to provide correct responses in complex auditory scenarios does not necessarily reflect the amount of LE exerted. Although both measures tend to vary with task difficulty, they represent distinct aspects of auditory processing. Our results further support this concept.

Altogether, these results validate hypotheses **H2.1a** and **H2.1b**.

5.2 The relationship between EFs and SI/LE

In a second part of our investigation, we examined the relationship between EF capabilities and SI and LE.

> **H2.2a - Behavior**
>
> Performance in executive functioning, especially inhibition, would be correlated with SI in SIN and SIS scenarios.

> **H2.2b - Behavior**
>
> Performance in executive functions, especially inhibition, would be correlated with LE in SIN and SIS scenarios.

Updating

> **Conceptual Clarification**
>
> Updating is a core executive function closely related to working memory, described by Miyake et al. (2000) as the "monitoring of working memory representation". It is therefore often studied within the scope of working memory. Although these two constructs are not identical, they are strongly intertwined. This distinction is not always made explicit in the literature, and it is important to note that findings attributed to working memory sometimes rely on experimental protocols that use tasks also employed to assess updating capacity. Also, updating has been characterized as one primary mechanism through which working memory contributes to psychological functioning (Carretti et al., 2005).

In SIN, we did not find the expected results showing a positive effect of high updating capabilities on SI and LE. In SIS however, SI positively correlated with updating performance, specifically the dual N-back task, for all TMRs. In other words, better performance in updating in the N-back task was associated with better

SI in SIS condition. A positive correlation between letter memory and SI at 3 dB TMR, showed a similar relation, but only in the most favorable condition, adding another argument in favor of such a link.

Using the Reading Span test to assess working memory and high and low semantic corpus for the listening task, Stenbäck et al. (2021) showed that higher working memory capacity is beneficial to achieve good speech understanding and reduced LE in SIN and SIS situations. Many studies investigating the relationship between working memory and listening in complex environments use the Reading Span (RSpan) test to assess working memory capacity (Füllgrabe and Rosen, 2016; Stenbäck et al., 2016). Such methodological differences with our work could partly explain differences between our findings (absence of correlation between the updating tasks performance and SI in SIS condition) and those reported in the literature regarding working memory, closely intertwined with updating. In a meta-analysis, Füllgrabe and Rosen (2016) suggested that, contrary to common assumptions, the RSpan test may not always be a reliable predictor of SI in complex environment performance for the type of population we studied in this project (young listeners with normal hearing). Instead, tasks that specifically target high-level cognitive processes such as inhibition, shifting, updating, or other domain-general mechanisms like processing speed may provide more accurate predictions of individual differences in SNR understanding. This interpretation aligns with our hypotheses, suggesting that examining specific EF, such as inhibition, in the current project could offer more insight into listening in complex environments.

Inhibition

In addition to updating, a significant correlation between SI and inhibition performance in the Stroop task was found, but only in the most adverse SIS condition (-9 dB TMR). A lower inhibitory cost in the Stroop task was associated with better SI. At other TMR levels, like in all SIN conditions, no correlations were found between SI and Stroop performance. Altogether, these results describe an interesting pattern across SIS and SIN. They show that participants with smaller inhibitory costs in the Stroop task, reflecting stronger inhibitory control, achieve better SI in adverse listening conditions when some IM is involved.

This result is not particularly surprising since inhibitory cost has already been shown to relate to listening in adverse conditions (Stenbäck et al., 2016), illustrating the role of inhibitory control as auditory scene complexity increases. However, these results were not consistently replicated in another study using EM and IM while measuring LE with the ESCU scale. Despite the proposed role of inhibition in speech understanding under difficult listening conditions, they found no relationship with subjective LE (Stenbäck et al., 2021). In these studies, inhibitory cost was measured using the Swedish version of the Hayling Task (Burgess and Shallice, 1996). In this task, participants complete highly predictable sentences, and inhibitory control is assessed in conditions where they must produce an unrelated word instead of the expected one. The Stroop task, however, remains debated as a measure of inhibitory control (Aron et al., 2014; M. J. Moore et al., 2024). In fact, inhibitory control occurs when a response reading to be said or given is stopped. In this context, inhibition involves suppressing an automatic reading response while reporting a vocal response. However, participants may adopt alternative strategies, such as focusing on a single letter or looking slightly above the word, to minimize interference. To limit this, participants in our study were explicitly instructed to look directly at the word presented. Nevertheless, we cannot exclude the possibility of varied strategies during the Stroop task.

The observed relationship between Stroop performance and SI illustrates the relevance of studying inhibition in speech processing in SIS. We thus argue that inhibition, as elicited in our Stroop task, is partly involved in SIS to help segregate the talkers. Indeed, in SIN, the process of stream segregation requires increasing cognitive resources as the SNR decreases (Pichora-Fuller et al., 2016). In a multi-talker situation, the mechanisms involved are even more complex. Auditory stream segregation involves first separating the concurrent talkers, then selecting the target talker among the competing voices. As suggested by Lanzilotti et al. (2022) and Stenbäck et al. (2016), this selection process likely recruits inhibitory control. Moreover, since the correlation emerged only in the most adverse SIS condition, it is plausible that inhibition becomes particularly critical when the SIS conditions become more challenging.

The close interaction between updating of working memory and inhibition may also indicate their joint contribution to LE and SI. Inhibitory control may become increasingly important as listening conditions deteriorate, when updating alone is no longer sufficient to support complex speech understanding. Moreover, inhibitory control may operate at different levels: at an external level, by suppressing masking talkers or background noise, and at an internal level, by inhibiting irrelevant or incorrect predictions automatically generated during speech processing (Rönnberg et al., 2013). The emergence of inhibitory control as a factor that influences SI in more complex conditions suggests that this relationship is not linear. These findings may provide valuable insights into performance in complex auditory situations, in contrast to updating, which is generally related to speech understanding in a broader way, regardless of the environment's complexity.

However, correlations between EF performance and subjective LE did not mirror those observed for SI. In adverse conditions, the observed higher LE ratings may reflect task difficulty reported by the participants without necessarily engaging general EFs. As mentioned earlier, this subjective assessment may lack some sensitivity. To investigate the relationship between inhibition and LE, which is central to this project, objective physiological measures are likely to be more informative.

SIN and SIS differences

The observed correlations with EFs differed between SIN and SIS. Specifically, expected correlations between updating, inhibition capabilities and SI in SIN were not replicated (Ellis and Rönnberg, 2014; Stenbäck et al., 2016). Here again, the cognitive demands associated with the masking type could explain the difference between SIN and SIS in their relationship with EFs. In SIS, the presence of both EM and IM may generally require higher-level cognitive functions than in SIN. For instance, it could also be proposed that SIS engages more domain-general cognitive resources. To complete this explanation, Stenbäck et al., 2021 explained that the relationship between working memory and subjective LE is especially illustrated within the presence of IM in comparison with EM only. Our data support the idea that conditions with pure EM are not related in the same way as conditions combining EM and IM.

We suggest that in SIS, both segregation and selection processes are strongly engaged. Updating is generally recruited when listening, as soon as the situation becomes slightly complex. However, when the task difficulty increases further, inhibition is additionally required to ensure the correct selection of the target talker while inhibiting the masker talker. Compared to SIN with speech-shaped noise, SIS requires more nuanced inhibitory control to distinguish between similar voices. In other words, in the presence of IM, cognitive processes are increasingly recruited as the listening situation becomes more challenging.

Regarding our data, we cannot confirm hypothesis **H2.2a** and say that EFs are beneficial for speech understanding in SIS and SIN. Nonetheless, the observed relationships between SI in SIS, updating dual N-back task, and Stroop performance under adverse conditions illustrate the relevance of studying inhibition in speech processing in SIS. In addition, these findings support the use of the Stroop task as a measure of inhibitory control for the next step of the project.

5.3 EEG correlates of LE - alpha rhythms

Time frequency results

The EEG results of this study suggest that spectral analyses alone may be insufficient to fully capture the mechanisms underlying speech processing. When examining alpha and theta activities throughout the speech signal, we observed differences in alpha power between SIN conditions. While interesting, this global difference does not provide sufficient information on its own. Because the speech signal extends over several seconds, averaging band power across all electrodes and a second-scaled window may hide meaningful spatial or temporal effects.

Therefore, in a second step of the EEG analysis, we examined the topographic distributions of alpha oscillatory activities. We focused specifically on the alpha band to address the complementary hypothesis **H3**:

> **H3**
>
> EEG alpha oscillation dynamics would be impacted by the difficulty of the auditory scene.

The time dimension is particularly important when considering speech comprehension, and time-frequency analyses may be more suitable to study alpha evolution along a speech signal in a more interesting way than spectral illustration.

Time-frequency analyses of our data revealed increased alpha oscillatory activity in the left temporal region during speech listening for both SIN and SIS conditions. These differences allowed us to select a region of interest in the left temporo-parietal area, consisting of electrodes T7, C5, C3, P7, P5, P3, TP7, CP5, and CP3, for the subsequent ERSP computation. Furthermore, ERSPs showed that alpha differences between difficulty levels peaked around 1200 ms after stimulus onset in both SIN and SIS scenarios. Based on this observation, it could be suggested that an increase in alpha power band is associated with task difficulty in complex auditory situations.

However, this interpretation does not align with a common one that stipulates that an increase in alpha power reflects idling activity of underlying brain areas.

Several explanations can be considered. First, as the task may be very difficult in some conditions and for some participants, this counterintuitive result might represent a disengagement effect of some of them. However, this interpretation can also be opposed, considering behavioral results showing SI fairly above the chance level in the most adverse conditions.

Second, the observed effect could reflect different sub-alpha components. Mu and tau rhythms for instance (see 2.2, Wisniewski et al., 2017) have been identified in the 8-12 Hz EEG band. It has been suggested that tau rhythms are sub-alpha temporo-parietal rhythms related to auditory processing, which could partly align with our findings. For this component, increasing alpha power might indicate synchronization rather than desynchronization. Consequently, and has already suggested (Jenson et al., 2015), the increase in alpha power could reflect inhibition of the distractor, which is discussed in the next section.

The diversity of alpha origins

> ### Conceptual Clarification
>
> In this project, we distinctly refer to cognitive inhibitory control as an EF on one hand, and neural inhibition, reflected by alpha event-related synchronization (ERS) on the other hand. Although these concepts share similar terminology, they represent distinct mechanisms. Cognitive inhibitory control is detailed in Section 1.1 of Chapter 3 and the inhibition hypothesis in Box 2.2 of Chapter 2.

The alpha activity is complex and results from various brain activities, making it difficult to define and interpret. It involves event-related synchronization (ERS) and event-related desynchronization (ERD) in task-related brain areas (Jensen and Mazaheri, 2010). The alpha power inhibition theory (see Intro Box ERS/ERD; Klimesch et al., 2007) hypothesizes that the augmentation of alpha, resulting from more synchronization of the oscillations in a region (ERS), is associated with neural inhibition. Conversely, a decrease in alpha power, resulting from desynchronization, is interpreted as increased neural activation. Based on this theory, it is commonly proposed that EEG alpha power can provide objective information about LE (Alhanbali et al., 2019). Furthermore, speech perception-related regions such as the IFG or the auditory cortex are considered potential sources of these alpha power changes. Also, the auditory-related alpha temporo-parietal sub-rhythms, known as tau rhythms (see Box 2.2 in Chapter 2) are thought to be particularly representative of listening processes in complex auditory situations (Lehtelä et al., 1997; Weisz et al., 2011; Wisniewski et al., 2024).

However, there is currently no clear consensus regarding the dynamics of alpha activity during complex listening, nor about its relationship with LE. In other words, there seems to be an agreement that no consensus exists (Alhanbali et al., 2019; Paul et al., 2021; Wisniewski and Zakrzewski, 2023). Hence, some studies report that alpha power increases as listening conditions become more difficult, whereas others observe the opposite pattern.

For instance, Obleser et al. (2012) found increased alpha power when participants listened in more adverse conditions. Similarly, Hall et al. (2019) reported that mean alpha power at the centro-parietal electrode CPz increased as the SNR decreased. Dimitrijevic et al. (2017) showed that attentive listening produced alpha ERS activity in some listeners, while passive listening showed almost no induced alpha oscillatory response. In the same line, Alhanbali et al., 2019 observed an increase in alpha activity in SIN. These findings are consistent with the view that alpha ERS reflects neural inhibition, whereas alpha ERD indicates neural activation

(Jenson, Thornton, et al., 2014).

In contrast, other studies have reported that alpha power decreases as listening becomes more complex or degraded. McMahon et al. (2016) observed greater alpha power when sentence intelligibility improved, suggesting that alpha suppression follows more difficult listening. Hunter (2020) found stronger alpha ERD, that is, a decrease in alpha power, during degraded and unpredictable sentences compared with predictable ones. Similarly, Seifi Ala et al. (2020) proposed that alpha ERS/ERD dynamics may better reflect speech performance than LE itself. They also reported a transition from ERD to ERS over prolonged listening, potentially indicating difficulty during speech understanding. Also, Wisniewski et al. (2021) showed stronger alpha suppression in parietal and occipital EEG data components extracted with ICA during filtered speech, corresponding to more difficult listening conditions.

Inconsistencies in alpha findings are common, and the literature remains unclear regarding the effect of speech complexity on alpha power. For example, Paul et al. (2021) and Dimitrijevic et al., 2017 reported different results concerning the relationship between alpha oscillations and effort ratings in left IFGs, although their experimental protocols were similar. Such inconsistent results do not necessarily indicate errors or bad experimental design, but rather emphasize that, especially in physiological and EEG research, factors such as data processing pipelines and analytical approaches can strongly influence the outcomes. Moreover, Alhanbali et al. (2019) suggested that inconsistencies across studies may also arise from differences in speech materials or noise types. In general, studies using different experimental protocols can show apparently opposing but potentially complementary results.

Furthermore, one major explanation could stand in the fact that multiple neural sources contribute to alpha activity (Klimesch et al., 2007; Wisniewski and Zakrzewski, 2023. Opposing behaviors of these sources during complex listening tasks may underlie the variability observed between studies. Such source differences can hide results and cause one particular rhythm to mask another when analyses are based solely on scalp topographies or time-frequency representations (Weisz et al., 2011), as we did in a first approach in this study. Therefore, authors suggest that the use of component-based analyses, combined with source localization, are essential for identifying the specific neural generators of alpha activity related to listening (Makeig et al., 1995; Wisniewski and Zakrzewski, 2023; Wisniewski et al., 2024).

Component analysis of EEG signal during listening showed different types of alpha activities. Wisniewski and Zakrzewski (2023) demonstrated that alpha power increased in some components while decreasing in others, and that this was related to different sources. In a left temporal cluster associated with a tau component, alpha oscillations decreased (reflecting stronger alpha suppression) during listening, while the mu-rhythm-associated cluster showed increased mu (subalpha rhythm) activity. Consequently, component-based analyses should complement time-frequency approaches, as certain effects visible at the component level may not be detectable in the overall time-frequency domain.

Thus, left temporo-parietal alpha increase revealed in our data during speech listening in SIN and SIS was refined with independent component analyses to investigate how independent components of these alpha dynamics are expressed depending on the difficulty level.

Independent Component Analyses

Consequently, the analyses were extended to better understand the time-frequency findings and determine whether they could reflect a mixture of different alpha sources. We performed ICA clustering analyses, following the recommendations of Wisniewski et al. (2024). For both SIS and SIN scenarios, we identified temporal and temporo-parietal IC clusters within the resulting clusters (see Figures 7.21, 7.20, and 7.22). The temporal IC clusters were similar to those described in the literature (Wisniewski and Zakrzewski, 2023), and their corresponding ERSPs showed alpha desynchronization, meaning a decreasing alpha over the speech signal. This suggests that these temporal IC clusters may correspond to tau rhythms. Similarly, the temporo-parietal IC clusters showed an increase of alpha activity across the speech signal (Figure 10.3 in Appendices), which could be related to mu rhythms. Taking into account the inhibition theory, this increase of alpha related to the mu IC clusters could be related to the diminution of somatomotor processing during listening (Ross et al., 2022; Wisniewski and Zakrzewski, 2023).

Our component analysis results help explain the apparent contradictions observed in the time–frequency domain. Thus, the increase in overall alpha power likely reflects a mixture of distinct alpha activities occurring during speech processing. Moreover, the left and right temporal IC clusters extracted from the SIS and SIN conditions, both showing alpha suppression, support the proposal by Wisniewski and Zakrzewski (2023) that alpha enhancement and suppression can co-occur during complex listening. These results also highlight that contradictory results in the literature concerning alpha during listening are, in fact, not contradictory, but complementary.

To further investigate and possibly confirm the tau and mu rhythms in our data, we computed exploratory source localization analyses of these IC clusters using Loreta software. The localized sources appeared to partially confirm the existing literature. Although we did not reveal significant contributions of the left IFG (see 2.5) some components appeared to have generators in the left temporal cortex (auditory) and in the left primary somatosensory and motor cortices. This is a supplementary argument in favor of possible tau (auditory) and mu (motor) sub-alpha contributions to SI in complex auditory situations. As these results only concern the SIN task, additional work will be necessary to clarify these findings.

5.4 Limits and Perspectives

This study has several limitations. First, concerning behavioral design, the Stroop task, used to measure inhibitory control, remains debated in the literature (M. J. Moore et al., 2024), and further studies could provide additional information concerning LE and inhibition by testing other inhibitory control tasks, such as the go/no-go or the Hayling tasks, for example. Also, only three SIN and SIS levels were tested, which limits the scope of the results. Future work should include a broader range of difficulty levels. Regarding the EEG data, the use of ICA analysis revealed interesting results regarding the potential use of alpha dynamics as an index for LE. However, source localization of the IC clusters should be analyzed again using different parameters (such as the number of ICs used for each cluster) to correspond to what is done in the literature.

Further analyses could extend the current findings in several directions. Considering the relationship observed in this study between Stroop performance and speech intelligibility in adverse SIS conditions, further analysis could investigate how this relation depends on the strategies listeners use in SIS. Specifically, examining whether the use of the auditory level difference between the target and masker talkers in negative TMRs could provide additional insights on how inhibitory control contributes to speech understanding in multi-talker, adverse listening situations.

Considering the EEG data, future analyses could extend the current findings in several directions, especially considering the density of EEG signal. Frontal midline theta oscillations, which reflect focused attention and have been suggested as potential indices of LE, could provide complementary insights into its modulation in complex auditory environments (Ishii et al., 1999; Onton et al., 2005). Considering the already existing results, more precise source localization of IC clusters would refine our understanding of the neural origins of the components identified in this study, and allow for the confirmation of if temporal IC clusters can be considered as tau-related IC clusters and temporo-parietal and mu-related IC clusters.

Additionally, a detailed comparison of tau and mu component alpha activation in relation to behavioral measures of SI and LE could clarify their respective contributions to auditory processing. Also, linking EF with IC clusters could allow for an integrated view connecting cognitive and neural measures in relation to LE and SI.

Many additional analyses are possible with the collected dataset, which remains available upon request, offering opportunities to further investigation of listening in complex auditory environments.

5.5 Conclusion

The behavioral results of this study report a relationship between the dual N-back updating task and SI in SIS conditions, and between the Stroop inhibition task and SI in the most adverse SIS condition. These results support the idea that updating and inhibition are both engaged in multi-talker situations. The absence of findings concerning the SIN conditions could be explained by methodological differences from the literature and the absence of IM which adds an additional dimension to the complexity of the auditory situation. These outcomes encouraged us to employ the Stroop task to train inhibitory control in the next phase of the research (see Chapter 8).

Concerning the electrophysiological results, we observed enhancement of time-frequency alpha dynamics in the left temporo-parietal region in both SIN and SIS conditions. Alpha-band EEG activity may reflects greater neural inhibition (Klimesch et al., 2007), which could indicate a change in listening processes during the task. This modulation seems linked to task difficulty and may also relate to LE. By refining these results by extracting independent components related to alpha activity described in the literature, we confirmed that alpha power observed in the time-frequency domain is the result of different independent components, and that both enhancement and suppression of alpha can be observed during listening in SIN and SIS.

It is also important to note that, in this study and others, the relationship with LE is not always explicitly addressed. Here, we examined how alpha dynamics evolve across conditions in SIN and SIS scenarios, which do not directly represent LE. In our data, alpha power increased under the most difficult listening condition of SIN scenario (-9 dB SNR, see Figure 7.13a), where behavioral results also indicated higher LE (Figure 7.6a). Although this does not establish a direct relationship between alpha power and LE, it suggests that such an association could be investigated in future work using the appropriate methodological rigor.

To demonstrate a direct relationship with LE, comparisons between subjective and objective measures would be required. However, as discussed previously (see Section 3.4), subjective and objective measures of LE rarely correlate, which completes the identification of its neural correlates. Inconsistencies across studies contribute to this challenge (McGarrigle et al., 2014). Nevertheless, this diversity also enriches the field, offering multiple perspectives and analytical approaches to address this complex concept. As often discussed (Alhanbali et al., 2019; Mohammadi et al., 2023; Paul et al., 2021), the question of whether alpha oscillations can serve as an objective measure of LE remains open. However, by continuing to refine analyses, particularly through independent component-based approaches, future research will help clarify this question.

Overall, these findings contribute to a better understanding of the neural mechanisms underlying listening in complex environments and highlight the need to disentangle the different alpha sub-components to interpret EEG results meaningfully.

8

Cognitive Training

Description of the Chapter

In a third study, the possibility of enhancing speech understanding in complex auditory environments through cognitive training is investigated. Building on previous findings that linked inhibitory control to speech intelligibility in adverse conditions, this study examines whether targeted inhibition training can lead to transfer effects on SI and LE in multi-talker situations. Additionally, we investigated the potential neural markers of LE related to alpha activity using EEG time–frequency and IC clustering analyses, to assess training-induced changes in these neural dynamics. This chapter aims to evaluate the transfer effects of cognitive training on speech understanding and how it may impact listening performance and its neural correlates.

Hypothesis 4

A cognitive training of inhibition would improve speech intelligibility and decrease listening effort in multi-talker situations.

Hypothesis 5

Cognitive training would impact alpha dynamics differently depending on the trained cognitive process.

Main Results

Inhibition-trained participants did not benefit from the inhibition cognitive training.

Therefore, no transfer effect of inhibition on CRM was possible.

Aligned to the previous study, alpha dynamics are resulting of multiple alpha components, including temporal and temporo-parietal components.

The left temporal component changed significantly pre- and post-training for the CRM-trained group.

1 Research Goal

In this study, we examined the effect of cognitive training on three different tasks on SI and subjective LE in SIS conditions. Three training groups were trained either on a SI task (CRM talker and masker), inhibition task (Stroop), or an active control task (CRM talker only). The goal of this study was to assess the following hypothesis:

> **Hypothesis 4**
>
> A cognitive training in inhibition improves speech intelligibility and decreases listening effort in multi-talker situations.

This hypothesis was based on the idea of transfer learning (see 1), and the results of the previous study (see Chapter 7) showing a correlation between SI and Stroop inhibitory task performance in adverse SIS conditions. More precisely, we hypothesized that inhibition control would be involved in SIS task to reduce masker interference. In multi-talker situations, suppressing irrelevant speech streams could allow the listener to better understand the target talker. Lower inhibition abilities may increase LE by making the listener try to process both target and masker streams simultaneously (Perrone-Bertolotti et al., 2017), whereas stronger inhibition capacities are generally associated with better speech intelligibility (Stenbäck et al., 2016). This suggests that better inhibition supports masking release and reduces LE. Additionally, the conclusions of previous work in our laboratory (Lanzilotti et al., 2022) suggest that some listeners rely on sound level differences as a segregation strategy and that this strategy may be based on inhibitory control.

Inhibition training

Previous research has explored inhibition training and its neural correlates. They reported diverse results and used different experimental protocols. Most of the studies trained inhibitory control using the stop-signal or the go/no-go tasks (Berkman et al., 2014; Chavan et al., 2015; Enge et al., 2014; Hartmann et al., 2016; Lenartowicz et al., 2011; Manuel et al., 2013). For instance, Chavan et al. (2015) trained participants in a go/no-go task over two weeks and observed changes in IFG activity, suggesting modulation of top-down inhibitory processes. Similarly, Berkman et al. (2014) used a stop-signal task over 20 days and found decreased IFG activation during inhibition execution but increased activity during the preparatory phase, suggesting a functional reorganization of inhibitory control networks. Furthermore, Lenartowicz et al. (2011) reported increased right IFG activation after two days of training with a visual stop-signal task. Complementary evidence from EEG work (Manuel et al., 2013) showed that one hour of auditory stop-signal training could modulate event-related potentials associated with IFG activity.

In our previous study, however, the only inhibitory task that significantly correlated with SI was the Stroop task. More precisely, a significant correlation was found between Stroop inhibitory cost and SI in adverse SIS condition. In addition, some other studies have also succeeded in training inhibition using the Stroop task (Davidson et al., 2003; Dotson et al., 2013).

As a consequence, we choose the Stroop task to train inhibition control, a motivation based on both our results (Chapter 7 , section 3.3) and the aforementioned literature.

Transfer effect

In addition to direct evaluation of inhibition control, the Stroop task has also been used to evaluate near or far transfer effects following inhibitory control or working memory (Enge et al., 2014; Karbach and Kray, 2009; Klingberg et al., 2005). Differences in training duration and consolidation time may also account for the variability in results across studies. For example, Enge et al. (2014) trained participants three times per week over three weeks, whereas Manuel et al. (2013) conducted only a single one-hour session, and Berkman et al. (2014) implemented a 20-day training protocol over three weeks. Such temporal differences could partly explain the discrepancies in outcomes, as the consolidation of learning inherently depends on the time factor.

Many inhibition training studies measure training outcomes without considering possible near or far transfer effects. Often, the training is considered effective when performance on the trained task itself improves, sometimes supported by neuroimaging evidence. Although this is informative, transfer effects of training have been demonstrated for other executive functions or working memory (Gathercole et al., 2016; Ingvalson et al., 2015). Concerning inhibition training transfer effects, Enge et al. (2014) compared adaptive training with active and passive control groups over three weeks of go/no-go and stop-signal training. Their goal was to show a near transfer effect of the training on the Stroop task and a far transfer effect on fluid intelligence. Although they observed a near transfer effect when comparing the trained group with the passive control, the absence of difference in the near transfer task (in this case, the Stroop task) between the trained group and the active control made them conclude that the improvement in the trained inhibition tasks could not be attributed to effective training of inhibitory control.

These results show the importance of the control group. Indeed, when comparing results between an active and a passive control, Enge et al. (2014) showed that the results differed; had they used a passive control only, the study's conclusion might have been different. Some studies do not use a control group at all (e.g., Lenartowicz et al., 2011), making it possible for the presented results to be due to general cognitive training more than task-specific training or even to placebo (Foroughi et al., 2016; see Box 1.3) or Hawthorne (Adair, n.d.) effects.

To address this main hypothesis, some other questions concerning the behavioral effects of the training were first addressed:

Complementary Hypotheses - Behavioral

H4.0: Training effect of inhibition: the inhibition-trained group would improve their performance in the inhibition tasks.

H4.1: Training would affect SI differently depending on the group and the TMR. Transfer effects would have a greater impact on SI in adverse conditions than in favorable ones.

H4.2: Training would affect the subjective LE depending on the group and the TMR. Transfer effects would have a greater impact on LE in adverse conditions than in favorable ones.

Finally, in the previous study, we observed an increase in alpha activity in the left temporal area, which may reflect an increase in LE. In the present study, we investigated how these hypothesized markers of LE would evolve during cognitive training. More precisely, we wanted to assess whether they could attest effects of different types of cognitive training at the brain activity level, addressing hypothesis 5 of the project :

> **Hypothesis 5**
>
> Cognitive training would impact alpha dynamics differently depending on the trained cognitive process.

As a consequence, we investigated the relationship between EEG activity and cognitive training by testing the following complementary hypotheses:

> **Complementary Hypothesis - EEG**
>
> **H5.1:** Reproduction of study 2: There is a difference in left temporo-parietal alpha activity between TMR in the pre-training session (regardless of the Group).
>
> **H5.2:** Behavioral group training effects (**H4.1** and **H4.2**) would be attested by alpha activity.

2 Material and Methods

2.1 Participants

Sixty participants were recruited for this experiment (19 women, 41 men, mean age 24.8 ± 5.02 years old). They all had normal hearing, as confirmed by pure-tone audiometry with an Elios® clinical audiometer (Echodia, Le Mazet-Saint-Voy, France), for the following frequencies: .25, .5, 1, 2, 4, 6, 8 and 12.5 kHz (hearing level $\leq$ 20 dB; mean hearing level 1.5 ± 10.5 dB HL, details in Appendix 10.1). Furthermore, participants had to meet the following eligibility criteria: no known hearing impairment, aged between 18 and 40 years old, native French language, no uncorrected visual impairment, no medication targeting the central nervous system, no neurological or psychiatric conditions, no known brain lesions, and no previous participation in a CRM task. They received €60 as financial compensation for their participation.

The participants were assigned to a group using a Latin square design to counterbalance order effects. They were unaware of the existence of multiple groups until the end of the experiment. This study was approved by the local ethics committee (IRB Number 2024 817). All participants provided written informed consent prior to data collection.

Sample Size calculation and recruitment

A power analysis using the *wp.rmanova* function from the *WebPower* package in R indicated that, for three groups and two repeated measures (pre- and post-training sessions), with a desired power of 90% and an effect size of .4, a total of 82 participants would be necessary. To account for an expected 10% data loss, we first planned to recruit 90 participants. Given the extensive protocol of this experiment, 60 participants were eventually recruited over a period of 10 months, corresponding to 300 experimental sessions.

2.2 Stimuli

Speech perception task

The SIS material from our French CRM corpus was used as the speech perception task. The task consisted of 2×5 TMR blocks of 16 trials each (-15, -12, -6, 0, and 6 dB TMR), presented in random order, along with an initial single-talker block. In each trial, the talker(s) were randomly chosen to minimize talker-specific effects. After each block, participants were asked to rate their subjective listening effort using the ESCU scale (see in Section 1.2 of the Chapter 5).

Inhibition task

An inhibition score was measured with a classic Stroop task and with an arrow ”Stroop-like” task. The inhibition task included four blocks (Stroop, Arrow-Stroop, Stroop, Arrow-Stroop), each comprising 82 trials.

Arrow Stroop : Direction names (”right”, ”left”, ”top”, ”bottom”), written in French (”droite”, ”gauche”, ”haut”, ”bas”), appeared one by one on the screen along with an arrow pointing to one of these directions (or a square as a neutral control condition). Participants were asked to indicate which direction was written in text while ignoring the arrow direction. Congruent trials presented an arrow pointing in the same direction as the text. Incongruent trials presented an arrow pointing in the opposite direction of the text (see Figure 8.1). For control trials, a square appeared instead of an arrow. This task was adapted from the arrow analog Stroop task from Shilling et al. (2002).

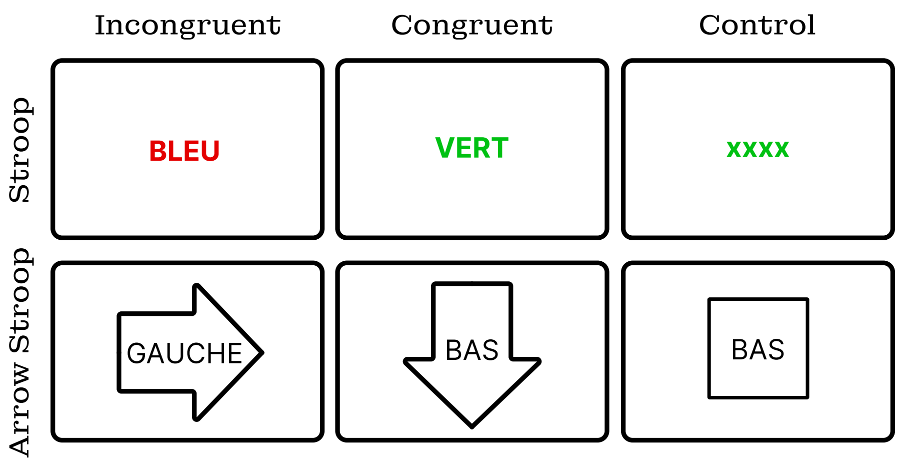


Figure 8.1: Example trials of congruent, incongruent, and control conditions for the Stroop and Arrow Stroop tasks. French task labels are translated as follows: ”Gauche” = left, ”Bas” = bottom, ”bleu” = blue, and ”vert” = green.

The Inquisit software used to present EF tasks in the previous study (see Chapter 7) does not support the inclusion of EEG triggers, which were essential for testing our electrophysiological hypotheses. Therefore, inhibition tasks were implemented in Python using the PsychoPy toolbox (Peirce, 2007) to allow precise synchronization with EEG recordings.

Control task

The control task aimed to reproduce the dependent variable task, in our case the CRM, without the associated difficulty regarding the measured cognitive ability (in our case SI and LE). Thus, we proposed a control task based on the CRM in a simple auditory situation, without noise or concurrent talkers. In order to keep the participants active, we asked the participants to indicate for each trial if a target call sign indicated at the start of a block was present for each trial.

2.3 Apparatus

Audiometry

The audition thresholds were measured with an Elios® clinical audiometer as described in the Section 1.3 of Chapter 5.

Audio and voice recording

As in the previous study, auditory stimuli were presented through ER-2 headphones (Etymotic Research), designed to minimize electromagnetic fields at the scalp, to avoid interference in EEG signal, at a sound level of 55 dB SPL, ensuring clear perception without discomfort.

For the Stroop task, the participant's verbal responses were recorded using a RODE (NT-USB Mini) microphone.

Electrophysiology

As in the previous study, EEG data were recorded using a 64-channel active system with a Biosemi Active II amplifier. Electrodes were installed using conductive gel (Signa gel) and placed according to the international 10-20 system.

2.4 Procedure

The cognitive training experimental protocol was conducted on 5 consecutive days. All participants began the first session on a Monday to ensure a regular interval between sessions.

The protocol consisted of a pre-training session on Day 1, followed by three training sessions on Days 2 to 4, and a post-training session on Day 5.

Participants were assigned to one of the three groups, each corresponding to a different type of training. They were not informed of the existence of the other groups. Participants performed the same tasks in the pre-training and post-training sessions. In the training session, they completed only the tasks specific to their group. The groups were defined as follows:

CRM-trained Group: Training on the speech perception task, using the CRM corpus in SIS condition.

Inhibition-trained Group: Training on the inhibition tasks.

Active control Group: Sham training, presented as a genuine training, using the control task.

At the beginning of each session, participants were equipped with an EEG cap and resting state activity was recorded following the same procedure as in the previous study (see Chapter 7).

Pre-training session (1h30) After providing informed consent, participants completed pure-tone audiometry. The general procedure of the experiment was then explained, without details about the cognitive training design. Then, the EEG cap was installed on the participant's head.

All participants, regardless of their group assignment, completed the same set of tasks. First, a 5-minute resting state was recorded. Participants then performed the speech intelligibility task (CRM). Finally, participants completed the inhibition tasks, with two repetitions of the Stroop and the Arrow-Stroop tasks.

Training sessions The training constituted of three sessions (one per day) of 45 min each. At the start of each training session, participants were equipped with the 64-electrode EEG cap, and a 5-minute resting-state was recorded. Then they performed the task according to their group:

Speech Intelligibility Group: Speech intelligibility task (CRM) in SIS condition. Single talker block (16 trials) + 2 x 5 TMR blocks (16 trials; -15, -12, -6, 0, 6 dB TMR in random order). Subjective LE assessed with the ESCU scale after each block.

Inhibition Group: Stroop and Arrow-Stroop task. 4 x 82 trials (Stroop, Arrow-Stroop, Stroop, Arrow-Stroop).

Control Group: CRM single-taker: 10 x 16 trials.

For all groups, each training session lasted between 30 and 45 minutes, depending on the participant's pace in completing the tasks and the time required to install the EEG cap.

Post-training session (1h30) The post-training session followed the same structure and order as the pre-training session: a 5-minute resting state, the speech intelligibility task, and the inhibition task. At the end of the experiment, the participants were informed about the study design and their group assignment and provided with a brief feedback upon request.

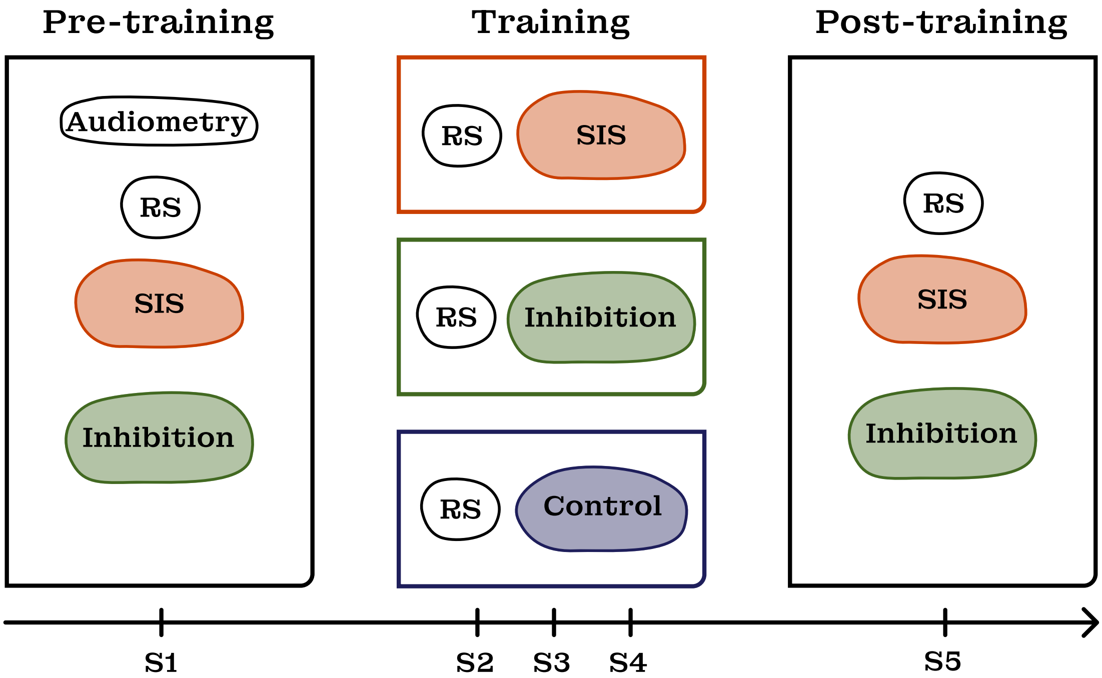


Figure 8.2: Cognitive training procedure over five consecutive days, from pre-training (S1) to post-training (S5) session.

2.5 Data analyses

The behavioral and electrophysiological data were analyzed using R 4.3.0 (statistics and behavioral plots), Matlab 2021a (EEG analyses), and Python 3.11.7 (speech-to-text Stroop extraction) and Loreta V20240713 (EEG source localization).

Behavioral analyses

Stroop The Stroop vocal responses were recorded with a microphone (RODE NTUSB mini) and saved as .waw files. The voice records were transcribed using the whisper speech-to-text model (OpenAI) implemented in Python, as described in Section 1.4 in the Chapter 5. The reaction time of each Stroop condition (congruent, incongruent, and control) were used for analyses.

CRM The percentage of correct responses was calculated as the number of trials in which both the color and digit were correctly identified, divided by the total number of trials in that experimental condition (16 trials in each block). Following current recommendations for speech analyses, scores were normalized using the Rationalized Arcsine Unit (RAU) transform in each condition (Studebaker, 1985).

Statistics Statistical analyses were carried out using MATLAB 2021a and R (version 4.3). A three-way analysis of variance (ANOVA) was conducted to examine the effect of TMR on SI and LE during training sessions.

Electroencephalography analyses

Outliers Participants P35 (group Inhibition), P36 (group Control), P53 (group Inhibition) were excluded for EEG data analysis due to poor data quality. In the time-frequency and ICs EEG analyses using the EEGLAB toolbox, if one session of a participant is unusable, the entire dataset for that participant must be removed to ensure consistent preprocessing and analysis across all sessions. In addition, to maintain balanced sample sizes between groups, three participants (P60 (group Control), P28 (Inhibition-trained group), and P31 (CRM-trained group)) were randomly excluded.

Preprocessing EEG data for spectral, time-frequency, and component analysis were preprocessed and processed as follows.

EEG data recorded during the CRM were extracted from the raw EEG files. Then they were average re-referenced, filtered on 4-25 Hz, and resampled at 512 Hz. Bad channels were removed and interpolated. Then, EEG data were split into data sets for each SIN and SIS condition, resulting in 9 data sets per participant. After that, mAMICA was applied on each condition data set, followed by independent component labeling with *ICLabel*. Eyes, muscles and artifact components were flagged in order to be excluded from the signal. Data were epoched on a window of -1000 ms to 2000 ms around the CRM sentences onset. More details on EEG pre-processing methods in Section 3.1 of Chapter 5.

Time Frequency For the time frequency analysis (see Section 2.4 in Chapter), one-way ANOVAs with permutations and FDR correction were applied in SIS (SNR) and SIS (TMR) conditions using the EEGlab STUDY statistics.

Independent Component analysis Component analysis allows additional exploratory observations of the EEG signal of the brain under SIS and SIN conditions. In this project, we aimed to reproduce the pipelines proposed in other studies (Jenson et al., 2015; Wisniewski et al., 2017, 2024). The component analysis (see Section 2.4) was performed on the component extracted with the mAmica algorithm, using the EEGlab study tool. For SIS and SIN separately, the component measures were precomputed, then a k-means clustering (with k = 13 and outlier clusters for ICs distant from more than 3 SDs from the centroid) was applied using time-based information (spectra and ERSPs) and location-based information (scalp maps). This clustering resulted in 13 components, which were analyzed per TMR or SNR condition, using one-way ANOVAs with permutation and FDR correction. Outliers were excluded from statistical analyses. However, they were included in the components and clustering computations.

Source localization From the mAmica results, the source localization (see Section 2.4) of all ICs of clusters of interest, revealed with the independent component analysis, were estimated using the Loreta software.

Microstates Microstates (see Section 2.4) metrics of each participant were computed using the tool described in Section 3.2 in Chapter 5. Resting state EEG recordings were extracted from the raw EEG files. Then a re-reference was applied on Cz, high-pass (1 Hz), notch (50 Hz) filters, and resampling at 512 Hz. Then the

MS pipeline was applied on participant and group levels (details in general methods). For clustering, k=7 clusters were selected, following Tarailis et al., 2024 recommendations. Microstates results are presented in the Appendix 3.5.

3 Behavioral Results

3.1 Inhibition Training

H4.0 - Behavior

Training effect of inhibition: the inhibition-trained group would improve their performance in the inhibition tasks.

In a first stage, results concerning within task inhibition training are presented to determine whether the inhibitory control training was effective on the Stroop performance. For this reason, only the data from the Inhibition and Control groups are included in this analysis.

Stroop

The three-way ANOVA results on Stroop conditions with Group (Inhibition-trained *vs.* control) × Session (pre *vs.* post-training) × Condition (congruent *vs.* incongruent *vs.* control) are presented on Table 8.1. No significant $group \times session \times condition$ interaction has been found. In addition, there is an absence of significant interaction $group \times session$. These results show that reaction times in the Stroop task were not significantly impacted by the training differently between groups (see Figure 8.3).

Df: degrees of freedom; F: F-test value; η_p^2: partial eta squared
Bold *p-values* are significant

	Df	F	η_p^2	*p-value*
Group	1	11.02	.23	**.002**
Session	1	9.95	.21	**.003**
Group:Session	1	3.60	.09	.0657
Condition	2	205.60	.85	<**.001**
Group:Condition	2	6.47	.15	**.005**
Session:Condition	2	8.41	.19	**.001**
Group:Session:Condition	2	.58	.02	.54

Table 8.1: Three-way ANOVA on reaction times, with Stroop Condition (congruent, incongruent, control), Session (pre- vs. post-training), and Group (inhibition-trained vs. control), as factors.

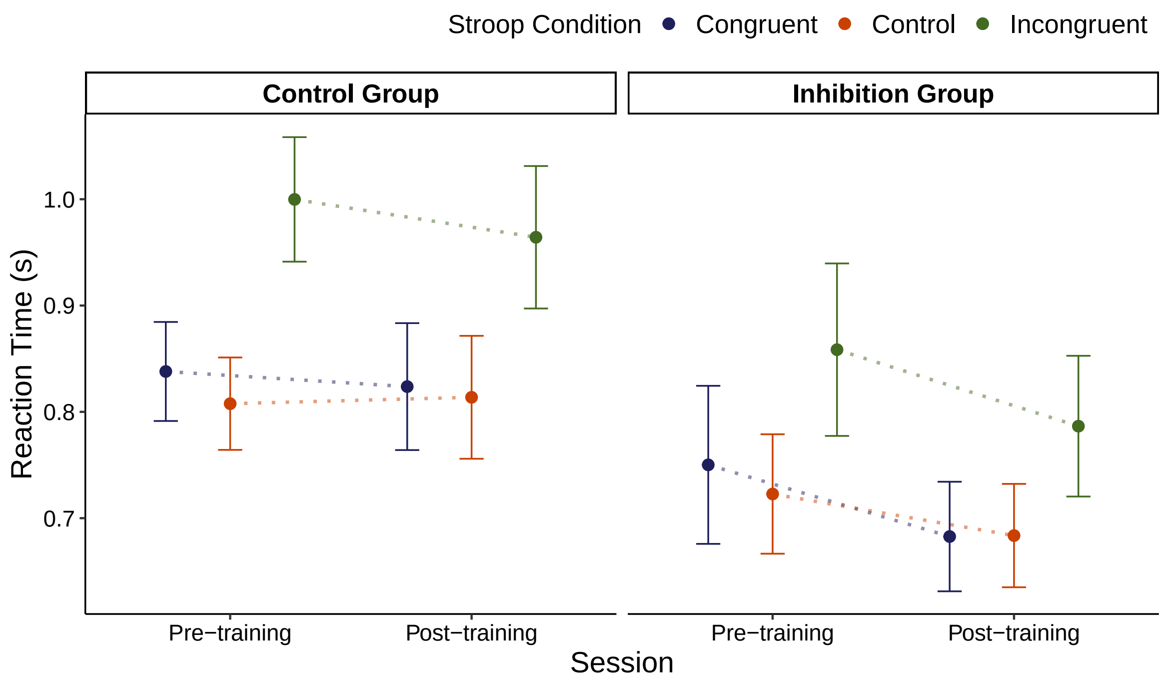


Figure 8.3: Mean Stroop reaction times (s) with 95% confidence intervals for each Stroop condition (congruent, incongruent, control), in control and inhibition-trained groups, across pre- and post-training sessions.

Arrow Stroop

The three-way ANOVA results on Arrow Stroop conditions with Group (Inhibition-trained *vs.* control) $\times$ Session (pre *vs.* post-training) $\times$ Condition (congruent *vs.* incongruent *vs.* control) are presented on Table 8.2. No significant $group \times session \times condition$ interaction has been found. However, the $group \times session$ significant interaction $F(1, 38) = 6.69$ $p = .014$, $\eta^2 = .81$) shows that the inhibition-trained and the control group were not affected similarly by the training. Tukey's HSD post-hoc tests on this interaction revealed a significant difference between the pre- and post-training sessions for both the control ($p < .01$) and the inhibition-trained ($p = .011$) groups. Both groups increased their performance in Arrow Stroop after training (see Figure 8.4).

Df: degrees of freedom; F: F-test value; η_p^2: partial eta squared
Bold *p-values* are significant

	Df	F	η_p^2	*p-value*
Group	1	2.26	.06	.14
Session	1	65.13	.63	<**.001**
Group:Session	1	6.69	.15	**.014**
Condition	2	165.82	.81	<**.001**
Group:Condition	2	.32	.01	.67
Ses:Condition	2	4.39	.10	**.018**
Group:Session:Condition	2	.06	.00	.93

Table 8.2: Three-way ANOVA on reaction times, with Arrow Stroop Condition (congruent, incongruent, control), Session (pre- vs. post-training), and Group (inhibition-trained vs. control), as factors.

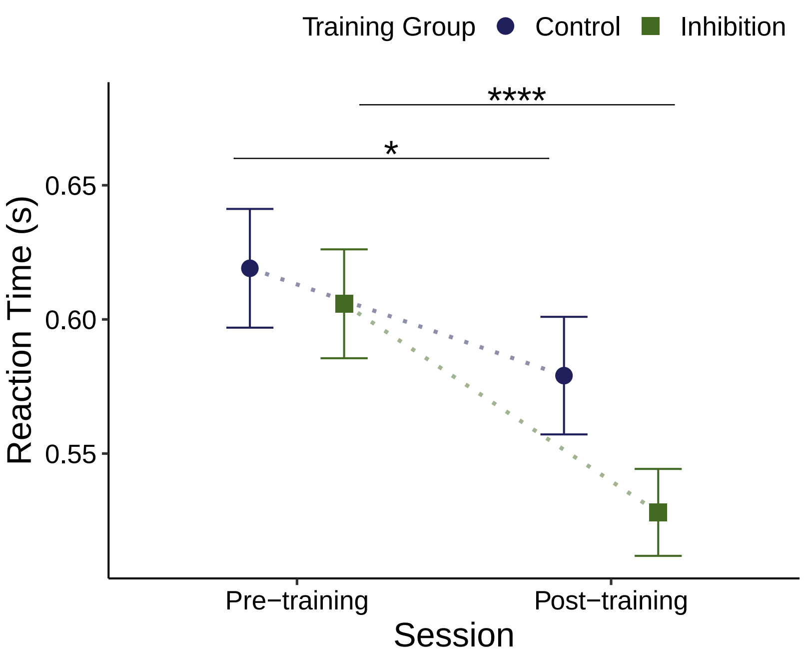


Figure 8.4: Mean Arrow Stroop reaction times (s) with 95% confidence intervals for each Arrow Stroop condition (congruent, incongruent, control), in control and inhibition-trained groups, across pre- and post-training sessions.

3.2 Transfer learning on CRM task

Speech Intelligibility

In the following figures, results are presented as the percentage of correct responses for clarity; however, all statistical analyses were conducted on RAU-transformed data.

H4.1: Behavior

Training would affect SI differently depending on the group and the TMR. Transfer effects would have a greater impact on SI in adverse conditions than in favorable ones.

The figure 8.5 presents the SI in SIS in % of correct responses of each TMR (-15, -12, -6, 0, 6) and all groups (SI, Inhibition, and Control) for the pre- and post-training sessions. The Figure 10.10 in the appendices presents the training results regarding each session and each TMRs for the three groups.

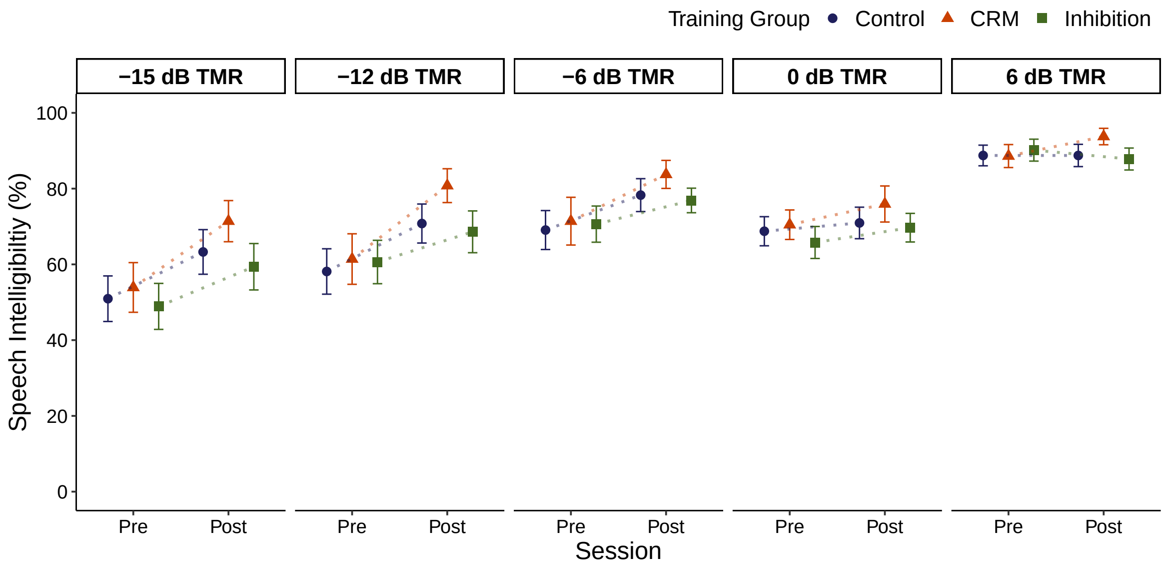


Figure 8.5: Mean speech intelligibility (% correct responses) with 95% confidence intervals for the three training groups (Control: blue; Inhibition: green; CRM: orange) across each TMR in the pre- and post-training sessions.

A three-way ANOVA on SI showed a main effect of TMR ($F(4, 131) = 136.7,\ p < .001,\ \eta^2 = .71$) and Session ($F(1, 57) = 88,\ p < .001,\ \eta^2 = .61$) corresponding to better scores for favorable TMRs and after the training. Significant interactions between Group and Session ($F(2, 57) = 7.34,\ p = .0015,\ \eta^2 = .20$) and between TMR and Session ($F(4, 220) = 14,\ p < .001,\ \eta^2 = .20$) showed that not all three groups benefited from the training. Tukey's HSD post-hoc tests revealed that this interaction was due to a significant positive training effect for both the CRM-trained and the control groups, but not for the inhibition-trained group. Furthermore, the CRM-trained group seemed to have benefited more than the control group from the training. Finally, no significant interaction between Group and TMR ($F(8, 132) = .31,\ p = .90,\ \eta^2 = .01$) nor between Group, TMR and Session ($F(8, 220) = .78,\ p = .62,\ \eta^2 = .003$) were found. The statistical results are detailed in Table 8.3 and Figure 8.6.

Df: degrees of freedom; F: F-test value; η_p^2: partial eta squared
bold *p-values* are significant

	Df	F	η_p^2	*p-value*
TMR	4	136.7	.71	**<.001**
Group	2	2.26	.07	.11
Group:TMR	8	.31	.01	.90
Session	1	88	.61	**<.001**
Group:Session	2	7.34	.20	**.0015**
TMR:Session	4	14.00	.20	**<.001**
Group:TMR:Session	8	.78	.03	.62

Table 8.3: Three-way ANOVA on Speech Intelligibility (RAU) with Group (CRM, inhibition and control), TMR, and Session (pre- and post-training), as factors.

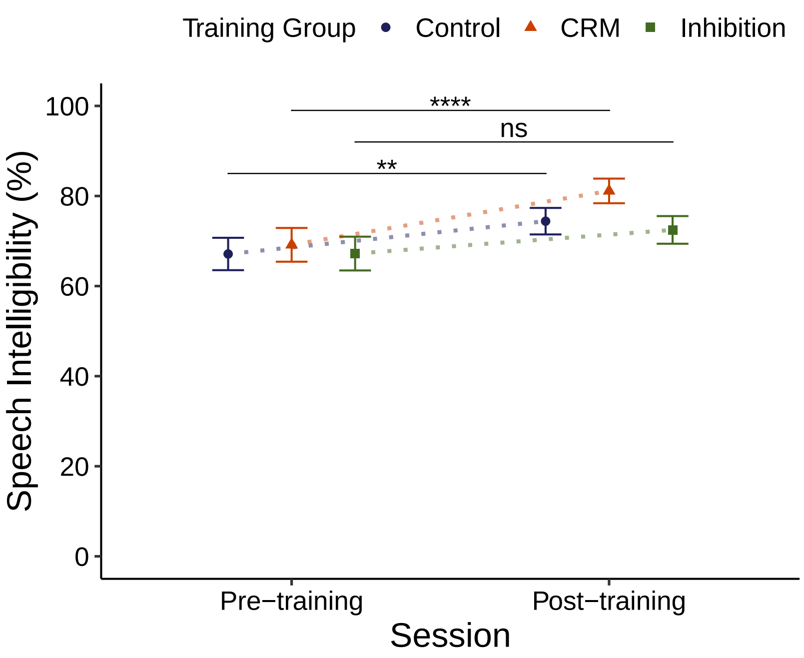


Figure 8.6: Mean speech intelligibility (% correct responses) with 95% confidence interval, across all TMRs in the pre- and post-training sessions, with p-values from Tukey's HSD tests (FDR-corrected) for the Session × Group interaction.

Listening Effort

H4.2: Behavior

Training would affect the subjective LE depending on the group and the TMR. Transfer effects would have a greater impact on LE in adverse conditions than in favorable ones.

The three-way ANOVA on LE is detailed in Table 8.4. The figure 8.7 presents the training effect on LE as assessed by ESCU scale for each TMR and each group.

Unlike the SI results (contribution 3.2), the three-way ANOVA showed a significant $Group \times TMR \times Session$ interaction on LE ($F(8,216) = 2.20$, $p = .03$, $\eta^2 = .07$) was observed. Tukey's HSD tests on this interaction showed that for the CRM-trained group a significant decrease of LE between was observed between pre- and post-training sessions at levels of -15, -12 and - 6 dB, while this was true only at -6 dB for the inhibition-trained group, and only at -15 dB for the control group (see Figure 8.8).

Df: degrees of freedom, F: F-test value, η_p^2: partial eta squared
Bold *p-values* are significant.

	Df	F	η_p^2	p
TMR	4	137	.71	**<.001**
Group	2	.06	.002	.94
Group:TMR	8	.65	.02	.65
Ses	1	16	.22	<.001
Group:Ses	2	3.24	.10	**.047**
TMR:Ses	4	7.49	.12	**<.001**
Group:TMR:Ses	8	2.20	.07	**.031**

Table 8.4: Three-way ANOVA on subjective listening effort (ESCU), with Group (CRM, inhibi- tion and control), TMR, and Session (pre- and post-training), as factors.

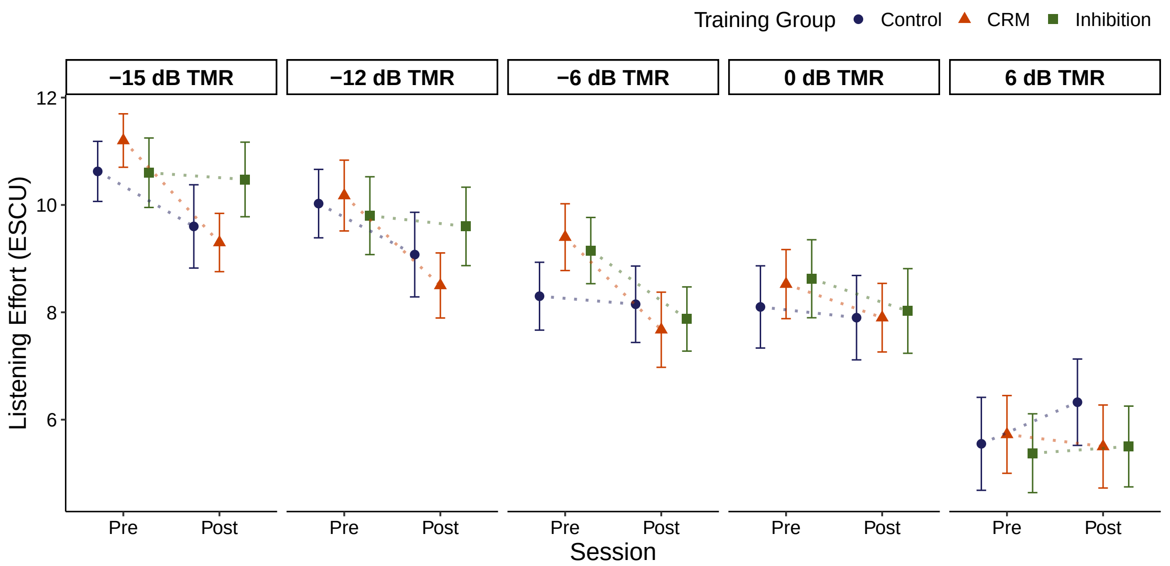


Figure 8.7: Mean subjective listening effort (ESCU) with 95% confidence intervals for the three training groups (Control: blue; Inhibition: green; CRM: orange) across each TMR in the pre- and post-training sessions.

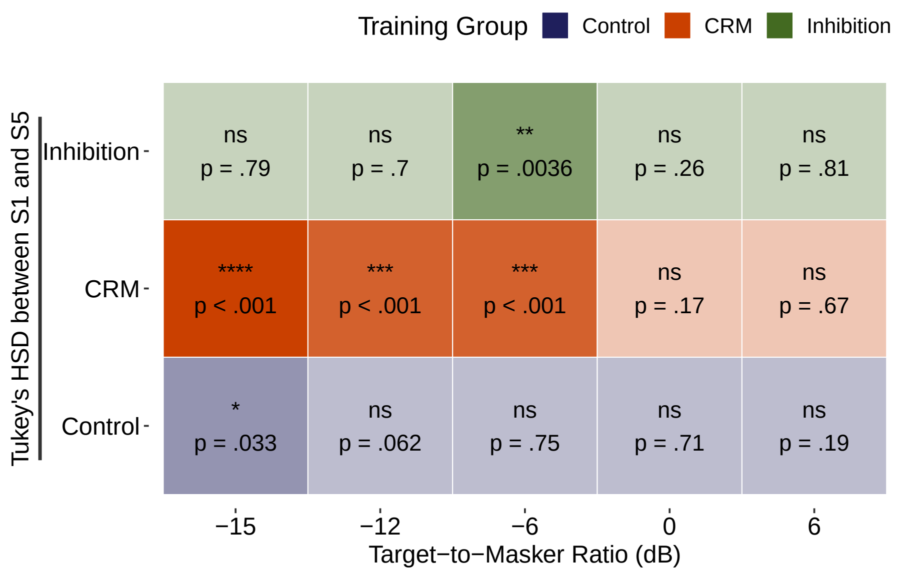


Figure 8.8: Tukey's HSD test results for training effects (pre- vs. post-training) on LE, shown for each group and TMR.

4 Electrophysiological Results

Hypothesis 4

Cognitive training would impact alpha dynamics differently depending on the trained cognitive process.

Complementary Hypothesis - EEG

H5.3: Reproduction of study 2: a difference in left temporo-parietal alpha activity between TMRs in the pre-training session would be observed (regardless of the Group).

H5.4: Behavioral group training effects (**H4.1** and **H4.2**) would be attested by alpha activity.

4.1 Time Frequency

Because behavioral results indicated a training effect in both the control and the CRM-trained groups only, and that the inhibition group did not improve in trained tasks (i.e., Stroop), EEG analyses were restricted to the CRM-trained and control groups.

Alpha power topographic activity

Pre-training Alpha activity before training was investigated to answer the following complementary hypothesis:

H5.3 - EEG

Reproduction of study 2: a difference in left temporo-parietal alpha activity between TMRs in the pre-training session would be observed (regardless of the Group).

Figure 8.9 illustrates the alpha-band activity (8-12 Hz) during the speech signal (0-2000 ms) in each TMRs of all participants in the pre-training session. FDR-corrected permutation tests revealed a significant difference in alpha activity in the left temporo-parietal area between the different TMRs. This difference corresponds to an increase in alpha power in a large part of the scalp as the TMR decreases. The details of topographic maps for each TMR of each group can be found in Appendices 10.12.

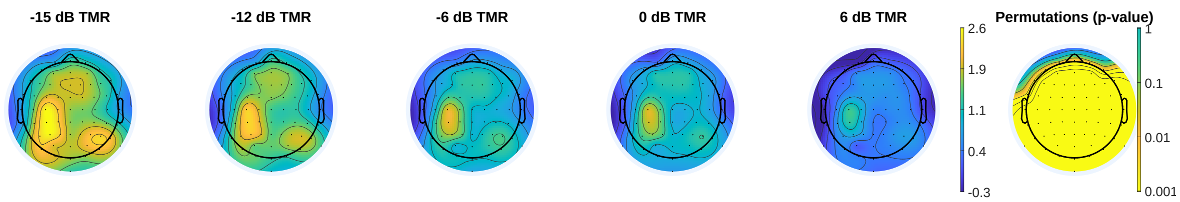


Figure 8.9: Topographical maps of mean alpha power (8–12 Hz; baseline-corrected) over the 0–2000 ms time window (0 = CRM sentence onset) in pre-training session. The right panel shows the p-value map indicating significant differences between TMRs. The topographies of each group can be found in Appendix 10.12

H5.4 - EEG

Behavioral group training effects (**H4.1** and **H4.2**) would be attested by alpha activity.

The Figure 8.10 illustrates the alpha activity (8-12 Hz) during the speech signal (0-2000 ms) in pre- and post-training sessions for the CRM-trained and control groups, regardless of TMR levels. FDR-corrected permutation tests revealed a training effect on the alpha power of the control group corresponding to higher values in centro-parietal areas, whereas no significant difference was observed for the CRM-trained group. In addition, no group effect was observed.

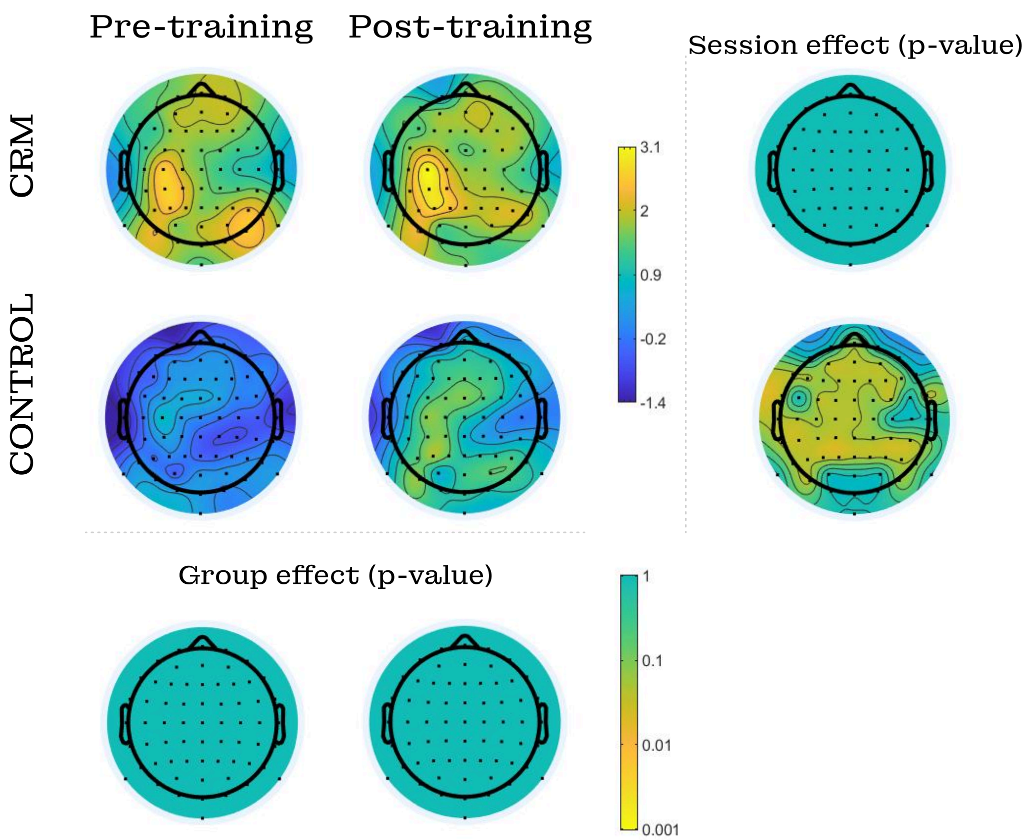


Figure 8.10: Topographical maps of mean alpha power (8–12 Hz; baseline-corrected) over the 0–2000 ms time window (0 = CRM sentence onset), showing the Session × Group (Control, CRM) interaction. The right and bottom panels display the p-value map of statistically significant differences, FDR-corrected.

Event-related spectral perturbation

The differences in topographic maps between TMR for the pre-training session show a significant difference in the majority of electrodes (see Figure 8.11). Thus, based on visual observation of the topographies and previous results (see Chapter 7), showing a significant difference on the left-temporal electrodes specifically, we selected the same electrodes T7, C5, C3, P7, P5, P3, TP7, CP5, and CP3, as a region of interest (ROI; see Figure 7.17) corresponding to the left temporo-parietal region, for subsequent ERSPs analysis. ERSPs were then computed on this ROI to compare the training effect between the CRM-trained and the control groups, regardless of the TMRs. The permutations tests revealed a significant group difference for theta band (4-8 Hz) and no difference at all between the sessions nor interaction between the group and the session.

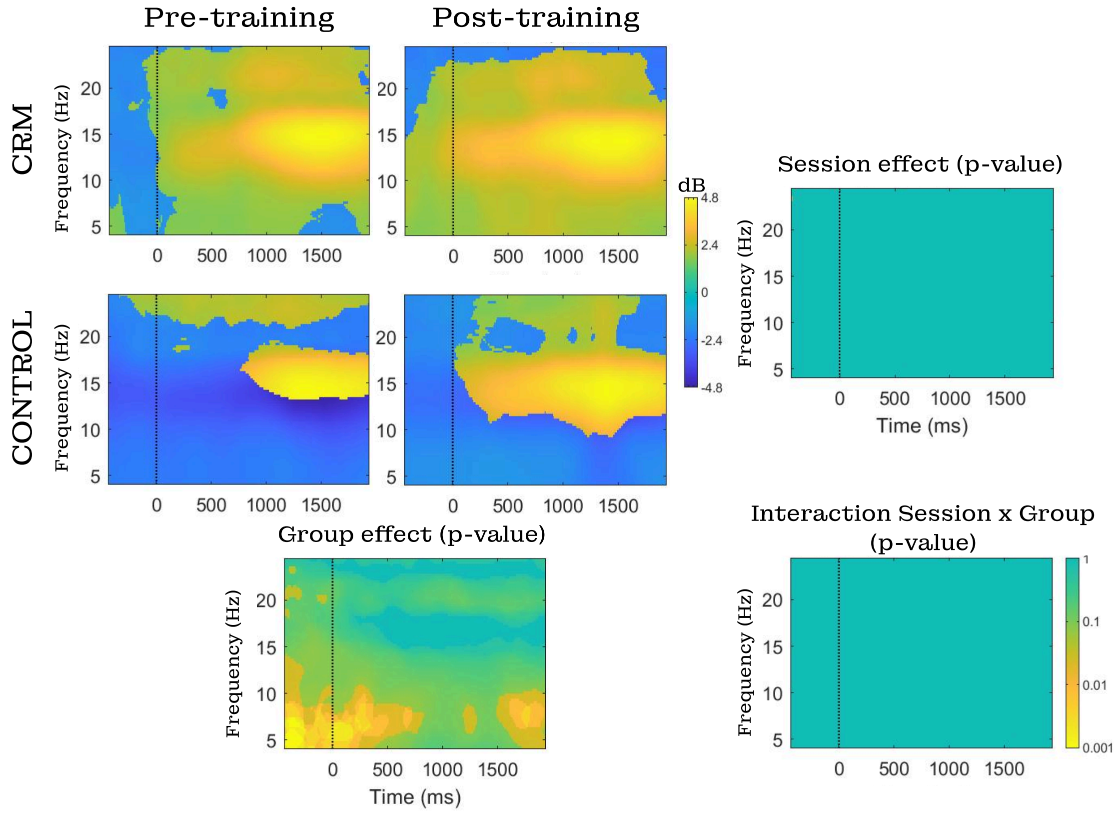


Figure 8.11: Time–frequency representations (ERSPs) showing the Session × Group (Control, CRM) interaction in the ROI (T7, C5, C3, P7, P5, P3, TP7, CP5, and CP3). The right and bottom panels display the p-value map of statistically significant differences, FDR-corrected.

Independent Component Analysis

> **H4.2b - EEG**
>
> Behavioral group training effects would be attested by alpha activity.

Left Temporal The left temporal IC cluster resulting from the IC analysis is presented in Figure 8.12. 379 components of 58 subjects were included in this IC cluster. On this Figure, the ERSPs associated with the left temporal component are represented for the pre- and post-training sessions of the CRM-trained and the control group. A significant increase in the alpha band was observed after training around 1500 ms after stimulus onset only for the CRML-trained group. In addition, a one-tailed t-test computed with sLoreta source localization software suggests that a main dipole generator is located in Brodmann area 6 (Frontal Lobe, Precentral Gyrus) for this left temporal IC cluster.

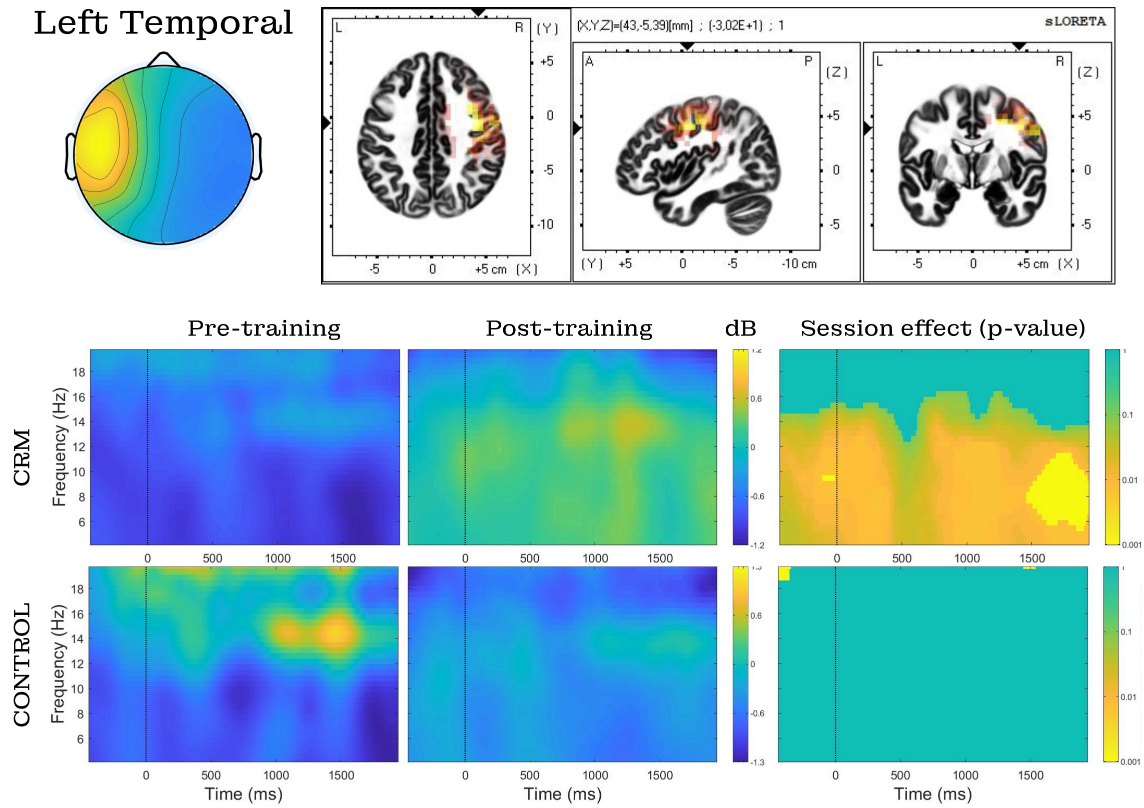


Figure 8.12: Average scalp maps of left temporal independent component clusters in the SIN condition. Left: average IC cluster scalp maps. Top-right: sLORETA source localization results. Bottom: ERSPs averaged for each pre- and post-training session of the control and CRM-trained groups.

Left temporo-parietal The left temporo-parietal IC cluster resulting from the IC clustering analysis is presented in Figure 8.13. 283 components of 59 subjects were included in this IC cluster (see Table 8.5). In this Figure, the ERSPs associated with the left temporal component are represented for the pre- and post-training sessions of the CRM-trained and the control group. In opposite to the previous component, no significant effect of training was observed in any investigated EEG frequencies. According to a one-tailed t-test, a main source of this component seems to be located in Brodmann area 3 (Parietal Lobe, Postcentral Gyrus).

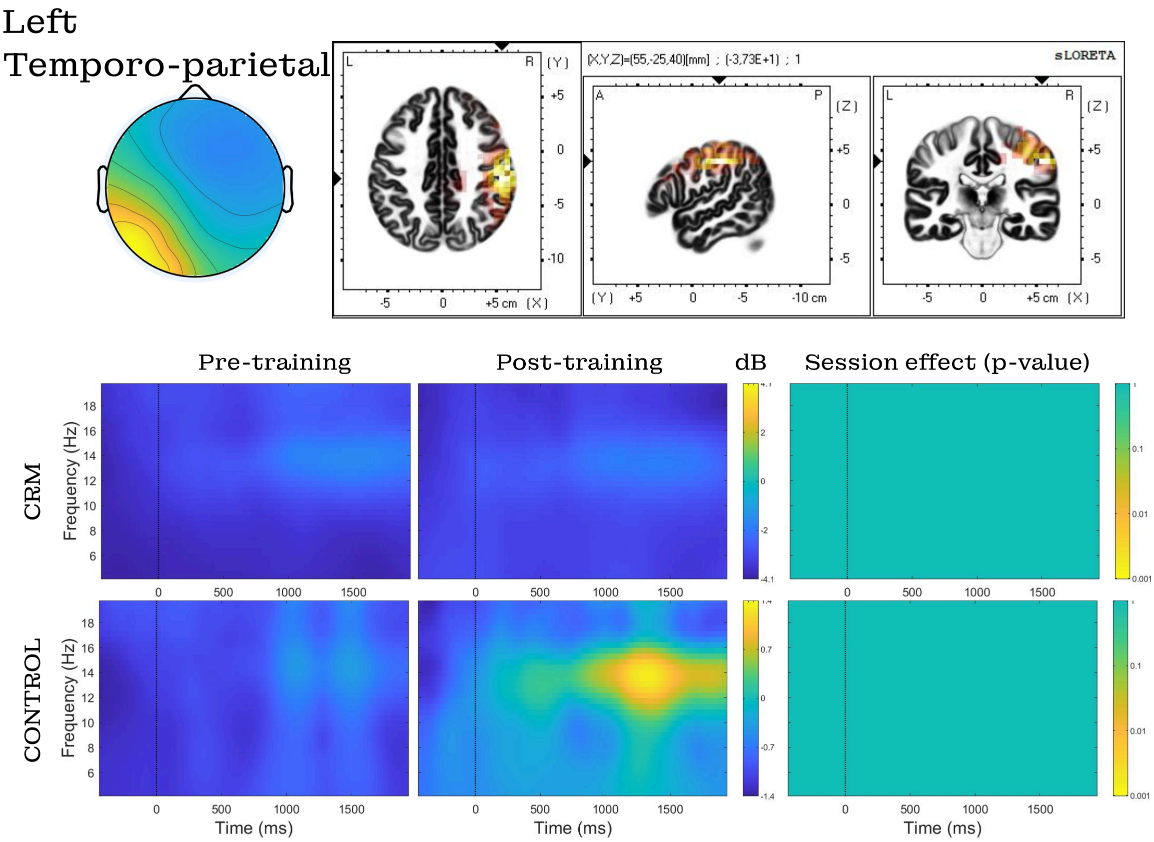


Figure 8.13: Average scalp maps of left temporo-parietal independent component clusters in the SIN condition. Left: average IC cluster scalp maps. Top-right: sLORETA source localization results. Bottom: ERSPs averaged for each pre- and post-training session of the control and CRM-trained groups.

Clusters composition

	Cluster		Subjects	ICs
TMR (n=60)	temporal	left	58	379
	temporo-parietal	left	59	283

Table 8.5: IC Cluster compositions with number of subjects, number of included ICs, and main cortical sources. BA: Brodmann area.

5 Discussion

Previous studies on inhibition have demonstrated training effect (Chavan et al., 2015; Davidson et al., 2003) as well as transfer effects to other cognitive tasks (Enge et al., 2014; Karbach and Kray, 2009). Moreover, inhibitory control is thought to play a key role in listening in complex auditory situations, especially in multi-talker situations (Lanzilotti et al., 2022; Stenbäck et al., 2016).

In multi-talker situations, we hypothesized that the level difference between the target and the masker talker serves as a cue for auditory stream segregation and separation and that inhibition of the irrelevant stream is essential for successful performance in adverse conditions (see previous study in the Chapter 7).

Building on these ideas, we asked whether inhibitory control could be trained to produce transfer effects on SI and LE in SIS conditions, as addressed by the following hypothesis:

> **Hypothesis 4**
>
> A cognitive training of inhibition would improve speech intelligibility and decrease listening effort in multi-talker situations.

Furthermore, considering that inhibitory control and speech understanding in complex environments share common brain areas within their neural pathways (such as the IFG, Lanzilotti et al., 2022; Zhang et al., 2017), and that alpha dynamics are described as a potential index for listening during effortful listening, we investigated the possibility of training-related changes in these dynamics, potentially varying depending on the trained task. Regarding this question, we addressed the following hypothesis:

> **Hypothesis 5**
>
> Cognitive training would impact alpha dynamics differently depending on the trained cognitive process.

5.1 The cognitive training of Inhibition

Absence of training effect in the Stroop tasks

In the present study, the comparison of the training effect between the inhibition and control groups did not lead to any interaction effect (see Table 8.4). More specifically, for the classical Stroop task, a group effect was present while no session effect was observed, suggesting a group inhomogeneity at pre-training session (see Table 8.1). This difference, however, was not present in the Arrow Stroop task (see Table 8.2). Instead, we found a main session effect, indicating improvements in both groups, which suggests an effect of familiarization with the task, as early as the second test session, rather than a specific effect of training in inhibition. The conclusion is that our training in the Stroop task and the Arrow Stroop task did not lead to conclusive improvement in inhibition.

This was surprising since previous studies have reported improvements in the Stroop task performance with practice (Davidson et al., 2003; Dotson et al., 2013). An important distinction lies in the training protocol: while our participants performed the inhibition task over five consecutive days, in two blocks of 82 trials each, participants in Davidson et al. (2003) completed only two sessions, but with a much denser design of twelve blocks of 128 trials per session, which may have resulted in larger training effects.

Taken together, these variations in the design and analysis of the experiment highlight that the metrics and protocols used to assess cognitive performance, in this case on the Stroop task, have a strong influence on the observed effects. Based on our analysis of inhibitory cost and the absence of interaction when comparing the inhibition-trained group with the control group throughout the sessions, our participants did not appear to benefit from the inhibitory control training regarding the Stroop task.

The absence of a larger improvement in the inhibition group suggests that Stroop tasks, as used in our training protocol, may not be optimal for training inhibition, as discussed earlier (see Discussion of Chapter 7). This limitation could partly explain the lack of transfer effect observed in both the SI and the subjective LE that we will discuss later. Despite our correlation results in the previous study and the results in the recent literature suggesting that Stroop tasks may be suitable for such a paradigm, it could be suggested that training effects on the selected inhibition task should be assessed prior to the experimental campaign, with a pilot study, for instance.

The Stroop task involves conflicts at both the stimulus and response levels (van Veen et al., 2001). At the stimulus level, participants must inhibit the dominant word-reading dimension to focus on the color-naming task. At the response level, they must suppress the prepotent verbal response associated with reading the word to produce the correct color response. This conflict, occurring at both levels, was an additional reason why Stroop was a candidate for the training of inhibitory control for an expected transfer on the CMR task.

Furthermore, Enge et al. (2014) also questioned whether inhibitory control can be effectively trained. In their study, although reaction times decreased in the trained task, the error rate increased, raising concerns about both the effectiveness of training and the level of participants' engagement. Due to time constraints, error rate analyses are planned for subsequent work. Compared with updating (or working memory) training, inhibition training appears more difficult to assess, in part because of potential changes in participants' strategies during the sessions. Similar strategy shifts have been reported in multi-talker listening training (Lanzilotti et al., 2022), where participants modified their approach to the task over sessions. This makes an additional similarity between inhibitory control and listening in SIS situations mechanisms.

Such changes in strategy raise the question whether the same cognitive processes are being engaged throughout training. It is possible that, by adopting new strategies, participants rely on different and complementary cognitive mechanisms than those used in the initial pre-training session, making it difficult to assess the targeted process.

5.2 Cognitive training in this study

> **H4.1: Behavior**
>
> Training would affect SI differently depending on the group and the TMR. Transfer effects would have a greater impact on SI in adverse conditions than in favorable ones.

In this study, our main hypothesis was that inhibition training would provide the participants with a better ability to segregate the masker from the talker. Indeed, according to Lanzilotti et al. (2022) and Stenbäck et al. (2016), it was hypothesized that inhibition would play an important role in SIS because of the need to isolate talkers and select the target talker's auditory stream. To test this hypothesis on SI or subjective LE, three-way ANOVAs were conducted, including group (CRM-trained *vs.*

Inhibition-trained *vs.* Control), TMR, and session (pre- *vs.* post-training) as factors. A main effect of Session indicated that, overall, participants improved performance in SI after training. The $TMR \times Session$ first-order interaction further revealed that this improvement varied across TMR levels, meaning that performance gains were not uniform across difficulty levels. More precisely, and consistent with previous findings in SIS (Chapter 6), a ceiling effect was observed in the most favorable condition (6 dB TMR), making further improvement at this level less possible than at more adverse levels. For the more adverse TMR, and as previously discussed (see Discussion in Chapter 6, Andeol et al., 2011; Lanzilotti, 2021), participants may adopt different strategies considering the SIS scenario differently. Some are immediately able to segregate the louder masker in negative TMRs, whereas others are not. Furthermore, Lanzilotti et al. (2022) showed that participants who were initially unable to apply such a segregation strategy during their first exposure to two-talker CRM stimuli could, after training, reach performance levels comparable to those who had used the strategy from the beginning, and that this difference was not significantly different between participants who received prior information about this strategy and participants who did not receive additional information. Lanzilotti et al. (2022) suggest that this strategy development may be based on inhibitory control. They explain that when the TMR decreases and the masker talker gets louder in comparison to the target talker, the segregation process can get easier for some participants, and that the ability to inhibit the masker may allow the participant to achieve proper intelligibility.

Concerning expected transfer learning on SI (**H4.1**), the absence of second-order interaction between the three factors suggests that the main effects of TMR and Session on SI did not differ between groups. Thus, the training type did not affect the performance variation across TMRs after the training. The absence of inhibition training effects discussed before and detailed in Section 3.1 may partly explain this non-significant second-order interaction. Indeed, as the inhibition was not trained, we cannot expect a transfer.

That being said, another interesting first-order interaction of $Group \times Session$ (Figure 8.6) showed that training type did influence the global SI. Tukey's HSD post-hoc tests revealed that both the CRM-trained and the control groups did significantly improve SI after training, while the inhibition-trained group did not. However, it seems that the improvement for the CRM-trained group (mean progression of 12.0%) was larger than the one for the control group (mean progression of 7.3%). These results confirm that the training was more effective for the CRM-trained group, but that the transfer effect expected on SI after inhibition training did not occur, which obviously accounted for a lack of improvement in the trained Stroop tasks.

H4.2: Behavior

Training would affect the subjective LE depending on the group and the TMR. Transfer effects would have a greater impact on LE in adverse conditions than in favorable ones.

Regarding subjective LE, the release of subjective LE reported by the participants was not the same through TMR levels and groups, as shown by the significant second-order interaction of $TMR \times Group \times session$ (see Table 8.4).

The post-hoc results of training effect (see Figure 8.8) for each TMR revealed that the CRM-trained group reported significant training effects on subjective LE for negative TMRs. Meanwhile, the control group showed significant training effects only at the most adverse level (-15 dB TMR), and the inhibition-trained group only at -6 dB TMR. These results support the idea that subjective LE ratings may lack some sensitivity to accurately represent the exerted LE during SIS tasks. It is suggested that large variations in SI, like the one observed for the CRM-trained group, may be associated with changes in subjective feelings, but smaller effects, like those observed in SI for the control group, may not. In addition, we suggest that subjective responses, like those provided in the ESCU, can vary through sessions due to external or internal factors such as fatigue, environment, experience, or mood.

The control task

Looking at the factors that may have contributed to some improvements in the control group, we can argue that some of them may be related to the design of the control training. Our control task was designed following the recommendations outlined in 8, which specify that an active control task should use the same structure as the main task, but exclude the cognitive process of interest. In this study, we implemented a task based on the same speech corpus used to assess SI and subjective LE, the CRM corpus, with the target talker only.

To maintain participant engagement in this active control task, we asked them to identify the presence of a specific call sign for each trial. Although this task does not to segregate between talkers, a potential limitation is that memorizing the target call sign may have involved – and trained – working memory processes. We attempted to minimize this by displaying the target call sign on the screen throughout the task, but any tasks involving a response will recruit some working memory. This remains a valid concern since improvements in working memory related to the CRM material may explain some improvements in the two-talker CRM we assessed Gathercole et al., 2016; Ingvalson et al., 2015. Similarly, participants were repeatedly exposed to the same eight talkers across the three training sessions as those used in the pre-and post-training measures. This repetition may have introduced a familiarity bias. As explained by Johnsrude et al. (2013), familiarity with target and masker voices strongly influences SI in complex auditory environments such as multi-talker situations. Moreover, given that the CRM corpus contains highly constrained and predictable sentences, the limited vocabulary may have further reinforced this familiarity effect.

Thus, as described above, the control group did improve in SI after the active one-talker CRM task training. A possible experimental improvement would be to use a different speech corpus, or at least a version recorded by different talkers than those used in pre-and post-training sessions. Alternatively, the English version of the CRM corpus could have been used, providing both new voices and a change in language, which could have decrease familiarity and knowledge of the corpus on these two different aspects.

Although behavioral interpretations of training and transfer effects are crucial, exploring the neuroanatomical responses to these trainings may provide deeper insights into their underlying mechanisms. More precisely, it is important to investigate whether EEG metrics could explain, at least partly, the behavioral results to dispose of a more objective assessment than the subjective LE.

5.3 Electrophysiological correlates of cognitive training

Complementary Hypothesis - EEG

H5.1: Reproduction of study 2: There is a difference in left temporo-parietal alpha activity between TMR in the pre-training session (regardless of the Group).

H5.2: Behavioral group training effects (**H4.1**and **H4.2**) would be attested by alpha activity.

Previous studies showing neuro-physiological changes after inhibition training (Chavan et al., 2015; Manuel et al., 2013) or SI training using the CRM (Lanzilotti et al., 2022) suggest that cognitive training may impact the neural processes underlying these functions.

In the present study, the behavioral results indicated that inhibitory control training did not lead to observable training nor transfer effect, likely due to the experimental design choices discussed earlier. However, the training of the CRM-trained group was assessed and discussed, and reproduced previous results showing improvement of the CRM task (Lanzilotti et al., 2022), with the addition of a control group, consolidating these results. In the latter study, modification of the IFG activity as assessed by functional near infrared spectroscopy, was correlated with the improvement in SI in the CRM task. In an exploratory and descriptive perspective, and based on the EEG results of the previous study in Chapter 7), we chose to present the EEG results to provide an overview of possible neural dynamics associated with training in speech listening in SIS conditions.

First, an analysis was performed across the different TMRs of the pre-training session to assess replication of results of the previous study showing an increase in alpha power in the temporo-parietal region as TMR becomes more defavorable. Although similar patterns were observed, the statistical significance across the majority of electrodes did not allow us to confirm whether the difference was specifically localized in the temporo-parietal region for these data. Consistent with the findings of the previous study and according to the present results, the same ROI (see Figure 7.17 in Chapter 7) as in the previous study was therefore selected.

Thus, following the first-order interaction effect $Group \times Session$ (see behavioral results in Section 3.2), which revealed an overall increase in SI and a decrease in LE for the CRM-trained group when considering all TMRs, the EEG data were analyzed by combining all TMR conditions.

Given the results of the previous study showing an increase in temporo-parietal alpha activity with decreasing TMR in SIS, a decrease in alpha power was expected after training, with a stronger effect for the CRM-trained group compared to the control group. However, a difference between the pre- and post-training sessions was observed only for the control group. There was no significant interaction of Group x Session, no session effect for the CRM-trained group, and no group effects.

Time frequency A complementary approach using time-frequency representation (ERSPs) on the selected left temporo-parietal ROI showed an increase of alpha activity starting 1000 ms after the stimulus onset in the pre- and post-training sessions for both groups. However, the statistical analyses did not reveal any significant interaction between Group and Session nor a session effect, meaning that this metric was not sensitive to the observed behavioral effects. As discussed before (see Discussion of Chapter 7), time-frequency alpha activity is probably the result of a mixture of alpha from different sources that could explain this lack of sensitivity. To further investigate possible multiple contributions of alpha sources, we then expanded the analyses to ICs of the EEG signal.

ICs In the resulting clusters of this study, we found a left temporal cluster (Figure 8.12, which could be considered as left tau sub-alpha rhythm as previously suggested (Wisniewski and Zakrzewski, 2023). In addition, a left temporo-parietal cluster was observed (Figure 8.13) that could be considered as alpha activity related to mu sub-alpha rhythms. As discussed in the previous study (Chapter 7) and in accordance with the literature, alpha activity observed in the time-frequency domain may be related to multiple sources. In addition, tau rhythm characteristics are still elusive in EEG recordings (Wisniewski and Zakrzewski, 2023).

Nevertheless, examining training effects on the ERSPs of the hypothesized tau component of the CRM-trained and control groups revealed a significant difference in the 6 and 12 Hz EEG band for the CRM-trained group only, covering low alpha frequency around 1500 ms after speech onset (see Figure 8.12). These results suggest that the left temporal tau IC cluster may reflect a training-related effect specific to the CRM task. Furthermore, the difference observed shows that alpha activity in the post-training session is higher than in the pre-training one for the CRM-trained group. As discussed in Section 2.2 Chapter 2, the decrease in alpha power is associated with an increase in neural inhibition of cognitive processes, while an increase is more likely a release of neural inhibition. This means that higher recruitment of the related brain region occurred and, consequently, the associated cognitive process.

In the present study, we observed an increase in event-related synchronization of alpha rhythms in the post-training session associated with left temporal IC clusters of the CM-trained group. Along with behavioral results showing an improvement in SI and a decrease in LE, this electrophysiological result could correspond to a marker of the benefits of the CRM training. In other words, it is hypothesized that participants' improvements would be attested by increased event-related synchronizations of tau rhythms.

Regarding other component analyses, no significant training effect was observed for the left mu IC cluster. Despite a slight increase in alpha activity in the post-training session of the control group, this difference was not statistically supported and cannot be confirmed; probably due to some outliers in this group. Since mu rhythms are primarily associated with sensorimotor activity (Pfurtscheller et al., 1997), training a cognitive function such as SI may not influence these mechanisms and consequently the associated alpha rhythm. However, it should be noted that the ERSPs associated with the left mu IC cluster in the present study do not show the same ERS pattern starting around 1000 ms as observed in the previous study (see Discussion in Chapter 7), which had supported its identification as a mu sub-alpha rhythm. Moreover, it is possible that this component does not correspond to the left mu rhythm, but rather represents a different neural source or a potential false positive.

6 Limits

6.1 Sample size

It is well known that the sample size impacts the robustness of training and transfer effect results. Studies showing no transfer effect, like Teng and Poeppel (2020) work, tested more than 100 participants, while Berkman et al., 2014 included 60 and Lenartowicz et al., 2011 26 participants. In the present study, an *a priori* sample size calculation revealed that 90 participants would be necessary to obtain a sufficient statistical power to test our hypotheses. Unfortunately, only 60 participants (20 per group) have been acquired so far due to a lack of time in this project. One could argue that the completion of the 90 participants would have changed some results. Regarding the data distribution of the training effect in the Stroop task, however, it is unlikely that this effect would become significant with 10 extra participants in each group. Future studies could investigate whether training with other inhibition-related tasks, such as the stop-signal or go/no-go tasks, produces different or stronger transfer effects despite the absence of initial correlation between them and SI in our previous study. Inclusion of control groups within experimental protocols increases the required sample size. Training programs extending over several weeks or months are particularly challenging to implement due to potential participant dropout, changes in experimenters, and scheduling constraints. In particular, recruiting a representative sample of the general population can be difficult, as working adults are often less available for multiple training sessions within a week compared to students.

6.2 Inhibition Training

As discussed, because the inhibition training did not produce the expected improvement in Stroop performance, we cannot observe a transfer effect to the CRM task. Without effective training on the targeted executive function, the theoretical gain in inhibitory control could not occur. This means that no measurable transfer to LE or SI in SIS conditions could be observed.

Also, it is important to note that the control group in this study was active, but only with respect to the speech intelligibility task. Therefore, when comparing the inhibition-trained group with the control group, we cannot conclude that the control group functioned as an active control in the context of the inhibitory control training protocol. Nevertheless, even though the control task was not specifically designed to train inhibition, it is likely that by the second session, participants had already made some progress. This is supported by the absence of a Group × Session interaction when comparing post-training session of the control group with session 2 of the inhibition-trained group. The session effect shows overall improvement, but no significant difference between the two groups. These supplementary results can be found in Appendix 3.1.

6.3 ICs

The IC analysis was computed on the whole data set, meaning that the inhibitory-trained group data recordings were included. We computed the mAmica and the clustering before concluding on the behavioral results, and thus decided to present the results as they are now, considering that the statistical and ERSPs representation include only the CRM-trained and control groups.

Also, the IC analysis provides centroids (for instance, the tau component) representing the center of one of the clusters, composed of ICs from individual participants. However, each cluster is made up of a different number of ICs originating from different participants, leading to an unequal contribution of subjects and so groups within each cluster. With the current analysis tools we used (EEGLAB), it is not possible to statistically compare ERSPs or computed from the same cluster but containing different numbers of ICs.

To address this limitation, as Wisniewski et al. (2024) suggested, it could be a solution to use a single or a defined number of representative IC(s) per participant, selected according to a reliability criterion such as their GEV.

In the context of this project, the IC analyses are exploratory. Future work with more advanced expertise could focus on inter-group differences, particularly concerning the tau component, to determine whether its temporal or spectral characteristics evolve differently across groups following training.

6.4 Software

For the EEG analyses in this study, we used the MATLAB toolbox EEGLAB. A STUDY structure was create,d including Session, Group, and TMR as factors. However, the EEGLAB study plugin has a limitation regarding the number of factors that can be implemented for statistical analyses. Specifically, three-way ANOVAs cannot be performed directly within the basic study framework, and the use of the LIMO EEG plugin (C. Pernet et al., 2016; C. R. Pernet et al., 2019) is required for such analyses. In this study, due to time constraints, we were unable to implement these analyses. Nonetheless, conducting them in future work would be highly valuable to provide a more detailed understanding of the differences in IC cluster ERD and ERS across TMR levels, sessions, and groups.

6.5 Data size

For future studies involving large EEG datasets, we would recommend using computers with sufficient memory, RAM, and processing power. When computing multi-run mAMICA decompositions, the use of parallel processing toolboxes is essential, as the scripts may take several days to complete. It is also important to consider the amount of data generated when working with a large number of participants, as storage requirements can easily reach hundreds of gigabytes. Proper data management and storage solutions are therefore necessary, and sharing such large datasets may only be feasible through dedicated platforms that support high-capacity data transfer.

7 Perspective

Several perspectives could be explored to build on these present findings.

First, considering the methodology of cognitive training of inhibition, future studies could employ different inhibitory control tasks, such as the go/no-go of the stop-signal tasks, for example. Combining these paradigms could also allow researchers to observe behavioral differences throughout training and to identify distinct patterns of transfer effects of inhibition training.

As the goal of this training was to observe transfer effects on speech understanding, the use of an auditory Stroop task (Knight and Heinrich, 2017, 2019) could also be the center of a cognitive training.

As the main objective of this project was to assess transfer effects on speech understanding, future work could also use an auditory Stroop task (Knight and Heinrich, 2017, 2019), which may provide a more ecologically valid framework for investigating the relationship between inhibition and speech perception in complex auditory environments.

Further analyses could also be carried out to deepen these findings. For instance, calculating error rates in the Stroop task would allow the assessment of performance changes in inhibitory control after training.

Regarding EEG results, exploring the correlation between subjective LE and alpha dynamics, in particular within specific IC clusters, could provide new insights into the extent to which these neural measures reflect listening processes in effortful situations. The present findings suggest that this direction of research deserves further investigation.

Furthermore, the EEG data were recorded on the whole signal of all sessions: there is still a lot to explore concerning EF activity and CRM. Triggers of the Stroop and Arrow Stroop are available in the signal, meaning EEG data regarding inhibition tasks could also be explored.

Moreover, as EEG data were recorded across the full duration of all sessions, there remains considerable potential for exploring additional aspects of EF activity and CRM performances. Since triggers from both the Stroop and Arrow-Stroop tasks are available in the dataset, EEG activity related specifically to inhibitory control could also be examined. Finally, the analysis pipeline developed for this cognitive training study could be reused and refined in future experiments, particularly those designed to address the methodological limitations identified in the current work.

8 Conclusion

This study investigated the possibility of mitigating listening effort using transfer effects of a cognitive training targeting inhibition. Unexpectedly, the inhibition training did not lead to improved inhibition performances, making it impossible to observe any transfer effects to LE of SI. However, we did observe training effects of speech-related training, with increased SI and decreased LE. These behavioral results demonstrated that it is possible to mitigate LE under the conditions and with the corpus used in this study. Future studies could explore this further with improved experiment designs, for example, by including other inhibitory control tasks such as go/no-go or stop-signal tasks, to better assess potential transfer effects of inhibition training.

Regarding electroencephalography measures during listening, we observed modifications of the left temporal alpha-related component, with an increase of ERS after training for the group that received complex speech-related training. This result suggests that independent components, in this case alpha-related ICs, may provide patterns to track and investigate neural indices of LE.

IV

Discussion

1 Summary and main results

This thesis project aimed to investigate listening effort using both behavioral and electrophysiological approaches. In addition, it explored the possibility of mitigating LE and understanding how such changes could affect the underlying neural correlates.

To do so, we followed a structured path. The first step was to define listening effort and the elements surrounding it in the auditory scene (Chapter 1), followed by an exploration of the neuroscientific aspects of complex listening environments (Chapter 2). Also, to make connections between the different cognitive processes engaged in such situations, we examined high-level cognitive processes that are executive functions (Chapter 3). Finally, to investigate the potential reduction of LE and improvement of listening abilities in complex auditory scenes, we focused on cognitive training and possible transfer effects (Chapter 4).

Along this path, several key questions were raised, leading to the formulation of hypotheses at each step.

Questions and Hypotheses

Q1: Does the language used for a speech corpus influence listening in complex situations?

H1: The use of a native language would positively influence SI and LE, particularly in adverse situations.

Q2: Is there a relationship between listening in complex situations and inhibitory control?

H2: Executive functions performances (especially inhibition) would be correlated with speech intelligibility and with listening effort.

Q3: What are the neural correlates of listening in complex situations?

H3: Alpha oscillations dynamics would be impacted by the difficulty of the auditory scene.

Q4: Is it possible to improve listening in complex situations by training associated cognitive functions?

H4: A cognitive training of inhibition would improve SI and LE in multi-talker situations

Q5: What are the effects of cognitive training on the potential neural correlates of LE?

H5: Cognitive training would impact alpha dynamics differently depending on the trained cognitive process.

To validate these hypotheses and answer the research questions, three experimental phases were implemented.

In the first phase (Chapter 6), we evaluated the behavioral performance of 50 participants in both SIN and SIS environments, with different levels of difficulty.

This first step allowed us to validate hypothesis **H1** and answer question **Q1**, confirming that the use of a corpus in the participant's native language, in this case French, had a positive impact on LE and SI in both adverse SIN and SIS conditions.

In a second phase (Chapter 7), the behavioral performance of 30 participants in SIN and SIS environments was compared with their abilities on nine EF tasks. The results of this second study allowed us to answer the question **Q2** and partly validate the hypothesis **H2**, showing that certain updating tasks and SI positively correlated at different difficulty levels in SIS. Additionally, a positive correlation was observed between SI in the most SIS adverse condition and inhibitory cost measured with the Stroop task. Participants showing better inhibition performances also had better SI in adverse SIS.

Furthermore, this second study allowed the investigation of the electrophysiological correlates associated with different levels of difficulty in SIN and SIS, addressing question **Q3**. Using various analyses of EEG signals, we explored the hypothesis **H3** and concluded that alpha oscillations showing increased activity during listening in complex environments, particularly in the left temporo-parietal region, were associated with multiple sources of alpha activity, probably corresponding to sub-alpha's tau and mu rhythms.

Finally, a cognitive training (Chapter 8) was conducted with 60 participants divided into three groups corresponding to different training types: inhibition control using the Stroop task, listening in SIS conditions, and an active control group. This third study showed no transfer of inhibitory control training to listening in SIS, but the listening training itself improved SI while reducing LE. While failing to support the hypothesis **H4**, it partially answered the question **Q4**, indicating that improvement of SI in SIS and decrease of subjective LE is possible. Nevertheless, potential directions were identified for future investigations.

Focusing on the effect of SIS training to answer the question **Q5**, the results also complemented previous findings related to question **Q3**, confirming that alpha activity in the left temporo-parietal region results from multiple components associated with alpha oscillations. Additionally, addressing the hypothesis **H5**, the investigation of a left temporal IC cluster, supposed to be associated with the alpha sound-specific tau rhythms, showed increased event-related synchronizations after training for the group that received a complex listening-focused training.

Across these three experimental phases, numerous questions and points of interest arose. The following discussion will address the main aspects observed, outline study limitations, suggest possible improvements and extensions, propose potential future directions, and conclude with the key findings of this thesis.

2 Defining and Studying LE

When Studying LE, it is important to consider the auditory scene as a whole, since multiple factors can influence – and are associated with – LE (see Chapter 1). However, when addressing specific research questions, including too many factors modifying the auditory scene could dilute the information and hide the main questions. Therefore, it may be useful to narrow down the scope by controlling variables that are not central to the research question. Of course, controlling these parameters reduces the ecological validity of the situation and, consequently, its resemblance to real-world listening environments. Various methods exist to simulate and measure effortful listening situations, and they can reflect distinct aspects of LE.

2.1 Ecological validity of complex auditory scene

Listening situations in the laboratory differ from those encountered in real life for several reasons. In the laboratory, the auditory levels of the target and the masker are strictly controlled, whereas in everyday life, sound levels constantly fluctuate. Moreover, the experimental designs of this project included only one masker, in SIS or SIN separately, while real environments often contain both background noise and multiple concurrent talkers, localized from diverse sources. In addition, although binaural listening is maintained, the conditions did not reproduce realistic spatial environments in which talkers are physically located in space, providing essential localization cues that support speech intelligibility. Furthermore, in certain SIS conditions, the same voice was used as both target and masker, a configuration that rarely occurs in real life and was perceived as particularly difficult, especially at 0 dB TMR. Without spatial or intensity cues, segregating two identical voices is very difficult, particularly with a corpus composed of rhythmically similar sentences. Moreover, the CRM corpus, used as the speech material for all studies, is simple and non-predictive, limiting the use of contextual and semantic cues that play a major role in natural speech comprehension. Listening through headphones also alters the listening experience by removing the multi-modal aspect of speech perception, as listeners cannot see the talker's face or gestures. As a consequence, it is important to keep these experimental choices in mind while interpreting our findings.

Some other factors are not related to the concept of experimental research, like the Hawthorne effect, for instance (see 1.3): participants are aware that they are being observed and that their performance will be analyzed. In training studies, even when participants are told that improvement is not expected, they may consciously or unconsciously aim to perform better or answer differently across sessions. Finally, participants were financially compensated for their participation, which introduces an external motivation rarely present in everyday listening situations.

Potential biases arising from laboratory conditions must be taken into account when interpreting results and reporting key findings in the literature. Although it remains challenging with current tools to fully reproduce real-world ecological environments in research designs, it is not impossible (Dehais et al., 2019; Scannella et al., 2018).

In this project, a first experimental phase was developed in order to control for potential bias regarding the population. Thus, the outcomes of this study confirmed the hypothesis **H1**, and validated the fact that, for better scientific validity, the use of a native-language corpus is recommended.

2.2 Hearing impairment

In this thesis, the inclusion criteria strictly required participants to have no hearing impairment. All participants were tested using pure-tone audiometry to ensure hearing thresholds were below 20 dB HL. This makes our sample unrepresentative of the general population, as discussed in 5, a substantial proportion of people experience some degree of hearing impairment during their lifetime. Moreover, the restriction to participants under 40 years old further limits the representativeness of our sample. Therefore, the conclusions drawn from this work should be interpreted in the context of this specific population, although they remain relevant for clinical applications or comparative studies focusing on similar groups.

Also, as detailed in 5, hidden hearing loss (HHL) is often reflected by an increase in LE in complex auditory situations, and remains difficult to diagnose. Interventions aiming to reduce LE could therefore be of particular value for this population.

Based on the findings of study 3 (Chapter 8), it appears possible to reduce LE through training involving complex SIS situations. However, we were not able to demonstrate a similar effect using inhibition control training, despite the observed relation between inhibition and SI in SIS condition in study 2 (Chapter 7). These results, showing a lack of induced correlation between LE and inhibition, may have suggested that inhibition is more directly related to SI than to LE, and therefore, inhibition training might not induce transfer effects on LE. Yet, given the absence of a clear inhibitory training effect in our data, no clear conclusions can be drawn, and further investigations are required to clarify this point.

It could also be proposed to design auditory training protocols for individuals with HHL based on SIS paradigms. However, the long-term effects of such interventions were not evaluated. These interventions could concern individuals using or not using hearing devices and should therefore rely on precise diagnostic assessments. By combining physiological measures such as pupillometry and functional brain imaging in large-scale studies, it might become possible to establish objective indicators of LE and define a LE threshold, such as for audiometric measures, which could serve as a clinical reference.

2.3 The subjectiveness of self-reported LE

LE can be assessed using various tools, as described in Sections 3.1 and 3.3. In the three studies conducted in this thesis project, the same self-report rating scale (ESCU) was used. This scale is simple to administer and allows for a rapid and intuitive evaluation of perceived effort. Accordingly, participants were asked at each TMR or SNR level to report their perceived LE using this scale.

However, using such a scale has limits and can lead to unexpected and unexplained results. For example, in our third study, regarding the second-order interaction of LE in $Group \times Session \times TMR$, we observed a significant post-hoc effect between pre- and post-training sessions' subjective LE at -6 dB TMR for the inhibitory-trained group. The surrounding difficulty levels (-9 dB TMR and -3 dB TMR) showed no significant results. This could be explained by the fact that the ESCU scale is not continuous and highly subjective in interpretation. The difficulty levels were presented in random order, but participants who had started with the most difficult level may have answered differently to the ESCU than participants have started with one of the easier level, adding some variance in the group results. Such analyses could be interesting for understanding how subjective LE evolves with dynamics changes of auditory environmental conditions. For example, a slight increase in noise may be perceived as highly effortful when it follows a period of silence, whereas the same level of noise might be rated as less effortful after exposure to a very noisy environment

2.4 Disengagement and Motivation

LE corresponds to the listener's deliberate allocation of mental resources (Pichora-Fuller et al., 2016, see Section 1.2), meaning that the listener actively decides to listen to a specific target. Importantly, effort does not necessarily involve performance; a listener may exert high effort without achieving better understanding, and good understanding is not always associated with high LE.

The FUEL model (Pichora-Fuller et al., 2016) considers motivation as a key factor. However, it was not measured in our studies. Motivation is a complex construct to assess, for similar reasons as LE, it is subjective and internally driven (Shields et al., 2022). Disengagement (see 2.2) may also occur in the most adverse conditions of listening in SIS or SIN conditions. Nevertheless, the average performance in our study groups was fairly higher than the chance level, suggesting that, overall, participants remained engaged even in the most challenging TMR or SNR conditions.

3 Alpha dynamics in SIS and SIN

When analyzing EEG results from designs including multiple factors, there are various possible approaches. In the present project, several difficulty levels, represented by TMRs and SNRs, were included in the listening tasks. This allowed the EEG recording to be examined either as a function of TMR/SNR levels or independently of them. Considering the EEG signal from a global perspective, by averaging across all conditions, provides a general overview of the neural dynamics involved in complex listening. However, this approach risks diluting condition-specific information that might reveal finer differences in how the brain adapts to varying difficulty levels. In the third study (Chapter 8), an additional factor was introduced with the multiple sessions. For simplicity, only pre- and post-training activities were analyzed in this report, although the full dataset from all sessions is available for future analysis.

To test or refine hypothesis **H3**, it was important to compare alpha activity across different TMRs and SNRs.

First, regarding the time-frequency domain, both studies 2 and 3 showed that alpha dynamics in temporo-parietal regions depended on the difficulty level (see Results 4.2 and 4.1). These statistical differences alone cannot explain how alpha oscillations are involved in LE or SI in complex auditory situations. However, the two studies reproducing similar results, with a larger amount of TMRs in the second study, resulted in the conclusion that in 90 different participants, alpha activity in the temporo-parietal region was significantly impacted by the TMR, starting around 1000 ms after the CRM onset. As discussed in the Chapter 7, as alpha is related to neural inhibition of the associated area (Klimesch et al., 2007), we would have expected to observe more desynchronization of alpha, meaning a decrease of power, with increasing difficulty. However, the time-frequency results in both studies showed opposite results, with increase of alpha power in difficult listening situations.

Independent Components

In both studies, we followed Makeig (1993) and Wisniewski and Zakrzewski (2023), suggesting the use of IC analysis to look at the data from a different angle. The results confirmed that alpha-band dynamics arise from multiple neural sources. In particular, temporal and parieto-temporal IC clusters were identified, potentially corresponding to tau and mu subtypes of alpha activity within the motor and auditory cortices. In addition, when examining the ERSPs associated with these components, distinct patterns of alpha power modulation were observed in both studies, supporting the idea that these sub-alpha sources may reflect different cognitive or sensory processes engaged during complex listening.

We provide here another evidence that IC clusters that showed topographic left or right temporal activation and associated ERSP with desynchronization of alpha oscillations (i.e., a decrease of alpha power) could be related to tau rhythms described in the literature. Similarly, IC clusters that showed topographic left or right temporo-parietal activation and which associated ERSPs showed increased synchronization of alpha oscillations, represented by an increase of alpha power, could be related to mu rhythms described in the literature.

When observing these temporo-parietal and temporal IC clusters across the different SNR or TMR levels (see 4.3), we saw that there are some significant differences across difficulty levels. Regarding the literature, we would have expected that the temporal IC clusters, supposed to be related to tau rhythms, would have shown an increase of ERD, meaning a decrease of alpha power, with increasing difficulty, and in opposition, the temporo-parietal, supposed to be related to mu rhythms, would have shown an increase of ERS with increasing difficulty.

In the SIN scenario, small but significant differences were observed between the ERSPs of temporal and temporo-parietal IC clusters across SNRs (see Figure 10.3). Similar to the ERSP results associated with the time–frequency dynamics (see Figure 10.2), these differences occurred around 1000 ms after CRM onset and were localized in the alpha frequency band.

In the SIS scenario, a significant difference of power between TMRs was observed for the left temporal IC cluster only (see Figure 10.4), and surprisingly, these ERSPs showed an increase of alpha power activation, suggesting we might have over-interpreted this IC cluster as tau-related.

Concerning the training effects, pre-post comparisons across all TMRs showed us that a significant difference was present regarding the ERSPs associated with the left temporal IC cluster. This significant difference was present for the CRM-trained group but not for the control group. We could suggest that the repetition of an auditory task may have increased the supposed auditory-alpha activity in the brain area related to this IC cluster. No such difference was observed for the left temporo-parietal IC cluster. As we supposed, this last IC cluster was associated with mu somatomotor activity, the absence of difference is not surprising and even consolidates the suggestion we made considering the left temporal IC cluster.

Methods Following Wisniewski et al. (2024) recommendations, tau-like temporal and mu-like temporo-parietal IC clusters were identified in 98.5% of participants in our studies. However, regarding the associated localized sources, they note that ”ICs of the tau type are localized along or near the superior temporal plane”, which was not the case with all of our results of the source localization approach. Despite this, we observed the characteristic alpha-suppression during sound presentation for some IC clusters, such as the left temporal IC cluster in study 2 (Chapter 7), for example. This suggests we may have overlooked some important processing details.

Outliers were included in the ICA and clustering computations, which is not an optimal practice and should have been done differently. Outliers should be excluded before computing any processing or analysis. However, they were excluded from the statistical analyses.

Still, some differences remained in our pipeline. We did not perform dipole localization using dipfit, as we had planned to use Loreta for source localization.Additionally, for the clustering and ERSPs computing, we included all ICs, rather than selecting the single IC per subject with the lowest variance rank as suggested. These differences may have impacted the results. However, sources were extracted using one component per subject for each cluster. Future analyses will be improved using the suggested adaptation, with the goals of achieving results more consistent with the literature, particularly regarding source localization.

Although it cannot be confirmed that tau and mu rhythms were observed, the left temporo-parietal alpha power increase seen in both studies (see Figures 7.20 and 8.12), which varied significantly across difficulty levels, likely originates from multiple alpha generators (Lehtelä et al., 1997). These sources should be further investigated in future work. As suggested by previous research (Jenson et al., 2015; Wisniewski et al., 2024), IC decomposition and the corresponding ERSP analyses provide valuable insights into auditory alpha rhythms, revealing ERS and ERD patterns during speech listening.

Source Localization

In this project, we attempt to localize the source of IC clusters that we supposed were related to temporo-parietal and temporal alpha activity. In the second study, generators in the Brodmann areas 22 and 42 were found for respectively the left temporal and the left temporo-parietal IC clusters. These sources, located in the superior temporal gyrus, are well characterized as part of the auditory cortical pathways involved in speech perception and phonological retrieval. They are often referred to as Wernicke's area, although their exact functional role remains a matter of debate (Binder, 2015, 2017). However, in both studies, other located sources were unrelated to the expected left auditory cortex or even left IFG potential sources. These analyses were exploratory, and we suppose we could improve the methodology. Furthermore, the interpretation of source localization is uncertain for auditory alpha ICs, as discussed by Jenson et al. (2015).

3.1 Microstates

In this project, resting-state activity was recorded at the start of each EEG session. This means that in total, we recorded 330 5-minute resting-states, which will be included in a larger laboratory-wide database that already contains data from more than 500 participants. The goal of this database is to analyze the microstate dynamics at rest.

Regarding this project, we had no hypotheses on the relation between resting dynamics and listening in effortful situations. However, in exploratory analyses, we extracted and analyzed the microstates' topographic maps and metrics for the second and third studies. The spatio-temporal dynamics of microstates at rest showed no interesting relationship with LE and SI, nor modification after cognitive training.

The absence of correlation between MS metrics and SI or LE results in both studies suggests that resting state spatio-temporal dynamics of functional networks may not directly reflect these behavioral measures. Further analyses could explore the temporal sequences or even MS during listening tasks. No specific hypotheses were made regarding MS; the results were presented primarily to illustrate the potential of this approach and to suggest alternatives to traditional time- and time-frequency analyses.

Finally, on more fundamental and methodological aspects, we observed that microstate topographic prototypes are stable across healthy participants (Michel and

Koenig, 2018), a result that is consistent with the literature. This robust method deserves further attention for its potential to predict cognitive performance. In particular, microstate A is often described as the "auditory" microstate, and additional data and statistical analyses focusing on it could provide further confirmation or raise critical insights on this topic. Moreover, microstates and IC analyses share common methods, with the clustering of maps of interest (ICs of GFPs), and their relation could be a very interesting methodological discussion.

4 Multidimensionality of Listening effort

4.1 IM and EM imply different release strategies

In this project, SIN and SIS complex listening situations were considered, always involving two streams: a target one and a masker one. In SIN, the masking, consisting of speech-shaped noise, was purely energetic, whereas in SIS, the addition of a second talker resulted in a combination of energetic and informational maskings. This distinction is important in interpreting the results of our study. It is now well known that different speech masking types influence speech processing (Peelle, 2022)

In SIN, as the SNR decreases, SI declines and LE increases. In other words, as the situation becomes more complex, performance drops and effort rises. In SIS, however, the SI curves show a floor effect at negative TMRs, when the masker talker is louder than the target. In contrast, subjective LE increases progressively, similar to SIN, as TMR decreases.

This difference between SIN and SIS can be explained by the nature of the masking. The presence of additional IM in SIS is mainly responsible for this difference, and we repeatedly confirmed in this project that SI followed different patterns in SIN and SIS (Brungart, 2001a). In SIN, decreasing SNR increases the EM progressively. In SIS however, decreasing TMR creates a situation where listeners can adopt different strategies and can use clues to their advantage in the presence of IM. Some listeners have similar performance patterns in SIN and SIS, showing a decrease in SI as TMR decreases. Others seem to use the relative level difference between the target and the masker voices as a cue. At negative TMRs, when the masker is louder than the target, these listeners can use this level difference to segregate the speech streams and select the quieter voice to focus on. Using this strategy, SI can actually increase at negative TMRs. In contrast, at 0 dB TMR (when target and masker voices are equally loud) this strategy is no longer available, making this condition the most difficult one.

Considering these two strategies explains the floor effect observed in SI curves, particularly in study 1 (see Figure 6.3) and study 3 (see Figure 8.5), where more negative TMR levels were tested. The study 1 results on individual SI in SIS (see Figures 6.6 and 6.7) illustrate this difference: some participants naturally discovered and used the level-based strategy, while others did not.

These differences in strategies and effects of EM and IM illustrate how experimental conditions influence performances, and why it is essential to clearly define the methods used to address hypotheses regarding listening in effortful conditions.

In several ways, it is evident that listening in SIN and SIS conditions is not the same. It can be assumed that a combination of the two would reveal additional differences across the various components of complex listening, such as LE and SI, and their associated neural correlates.

4.2 SI and subjective LE are not the same

In complex auditory situations, the intelligibility of the targeted auditory stream may require varying degrees of effortful listening depending on the environment, the talkers, and the listeners' abilities. This distinction between LE and SI was consistently observed throughout this project. Whether considering speech understanding in SIN or SIS conditions, or the relationships with EFs, SI and subjective LE tend to show different patterns. Differences were even found in their relations with Microstates (see 2.3 and 2.3 in Appendix). Moreover, conclusive results often centered on SI. The high variability of reported LE, combined with the use of restricted scales (as discussed previously), may partly explain the limited conclusions regarding subjective LE and further support the need for objective assessment tools to measure LE.

4.3 General Domain

LE is a multidimensional construct; no single measure can fully capture it, and different complementary methods should be used Alhanbali et al., 2019. High-level cognitive functions, particularly EFs, play a crucial role in speech processing under challenging conditions (Rudner and Signoret, 2016). In complex auditory environments, such as multi-talker scenarios, inhibitory control allows listeners to suppress irrelevant auditory streams and focus on the target talker. In this project this relation was explored in order to propose further interventions aiming to decrease LE in multi-talker situations.

The Stroop task, often used to index inhibition, reflects conflicts at both the stimulus and response levels van Veen et al., 2001, similar to the dual demands placed on listeners during speech-in-speech situations. Also, neuroimaging evidence highlights the left IFG as a key region supporting domain-general cognitive control processes that resolve conflicts in both language and executive tasks, suggesting overlapping mechanisms between LE and inhibitory control or EFs Novick et al., 2010.

Furthermore, behavioral and neural measures of LE, including EEG alpha dynamics, provide complementary insights, as alpha synchronization and desynchronization have been associated with inhibitory processes and effortful listening. Taken together, these findings indicate that variations in EF, particularly inhibition and updating, may partly explain individual differences in LE and SI, emphasizing the interdependence of cognitive control and auditory perception in challenging listening conditions.

5 Electroencephalography

5.1 The EEG data

EEG data can be analyzed in many ways, leading to various ways to look at the data. It is crucial to choose analysis methods that align with the cognitive processes being studied. In our studies, participants listened to speech presented for a few seconds. For this reason, we focused on time-frequency analyses and, based on recent studies and suggestions (Wisniewski and Zakrzewski, 2023), employed ICA to separate meaningful components of the signal.

Research using EEG often faces challenges related to the comparison of analysis pipelines. From one study to another, methodological differences can be substantial, and even slight variations may lead to large discrepancies in results and in their interpretation. In 2024, EEG celebrated its 100th anniversary as a tool for measuring electrical brain activity. On this occasion, Mushtaq et al. (2024) surveyed EEG researchers about current practices and future needs. One of the requests concerned the standardization of preprocessing and analysis pipelines, showing the need for more consistency and reproducibility in EEG research.

In this project, we decided to focus our interest on alpha oscillations, knowing that other frequency bands are described as potentially related to LE and SI and would be the center of further investigations. For example, frontal midline theta is associated with working memory during listening in complex auditory situations (Wisniewski et al., 2018). Additionally, other techniques such as event-related components P300 and N100 (Obleser and Kotz, 2011) or neural tracking, illustrating the relation between the EEG signal and the speech envelope (Ershaid et al., 2024; Muncke et al., 2022), have shown some sensitivity to listening effort.

6 Limits

Across the studies of this project, several limitations emerge that should be considered when interpreting the results and planning future work.

6.1 Participants

In total, 141 participants were included in this project (63 women, 77 men, 1 other, mean age = 25.7 ± 4.5 years). Regarding age, the sample is not representative of the general population. A key inclusion criterion required participants to be younger than 40 years old. This restriction considerably reduced the eligible population but was necessary to prevent potential decreases in hearing sensitivity related to natural aging of the auditory system.

Sample size and participant characteristics can influence the robustness of the findings. In particular, study 3 included fewer participants (60) than initially planned (90), which limited the statistical power to detect potential transfer effects of inhibition training.

Another limitation concerns the potential inclusion of participants with HHL. In this project, individuals were included in the studies based on standard audiometric criteria (auditory threshold $<$ 20 dB), which do not detect HHL. Consequently, some participants with undiagnosed HHL may have been included, which could influence behavioral and electrophysiological measures in complex auditory situations. Individuals with HHL might exert greater effort in less challenging conditions compared to non-HHL participants, and may disengage more quickly in highly adverse situations. This limitation is partly unavoidable, reflecting real-world population variability, but future studies could benefit from incorporating diagnostic tools for HHL to better characterize participants or create specific subgroups depending on the research questions.

6.2 Cognitive Training

In this project, we tested the effects of cognitive training on the short term, and longer-term assessments could provide more generalizable conclusions with potential practical applications. It is also important to note that training focused on a single target task does not necessarily generalize to the broader cognitive function it aims to enhance. Also, individual differences further complicate outcomes, as participants vary in baseline cognitive abilities, strategies, and responsiveness to training, which can influence whether training effects are observed (Katz et al., 2021.

Furthermore, the effectiveness of cognitive training remains debated in the literature (Gobet and Sala, 2023; Simons et al., 2016) and should be interpreted with caution and scientific rigor. It is important to adhere to scientific conclusions, as misinterpretation or overgeneralization of results can lead to misunderstanding and inappropriate applications.

6.3 EEG data quality

EEG is challenging to analyze for a non-exhaustive list of reasons. Interferences are sources of problems regarding the recorded signal: participants can move, yawn, clench their jaws when concentrating, or scratch their faces. Electrode issues can also occur, such as broken connections or poor contact with the scalp, which sometimes produce unusable recordings. Unexpectedly, some sessions may have nearly perfect data, while others resemble recordings during a magnitude 7 earthquake. Furthermore, participants differ in head size, shape, hair type, and sensitivity to the gel syringe, making it essential to allocate sufficient time to cap installation. Subtle aspects of EEG setup, including gel type and temperature, cap sizes, how experienced the experimenter is, or even electrode cleaning techniques, are rarely reported or discussed in research articles, yet they can substantially influence data quality.

6.4 Methodological choices

In this project, several methodological aspects of the experimental designs were informed by the findings of previous studies of the project. The use of a French corpus for studies 2 and 3 was validated by the results of study 1, and the choice of the Stroop task as an inhibition training paradigm in study 3 was based on study 2. However, this approach was not without its limitations. The cognitive training based on the Stroop task did not produce the expected training effect, which would have allowed us to confirm or challenge hypothesis **H4** and answer question **Q4** of this project.

This supports that, beyond methodological validation, a thorough literature review is essential when designing experimental paradigms. As discussed in study 3, inhibitory control can be trained using tasks such as the stop-signal or go/no-go paradigms, while the use of the Stroop task has produced more controversial results.

EEG analyses across studies were also constrained. While ICA decomposition and time-frequency analyses revealed promising insights, limitations in IC clustering, unequal IC contributions across participants, and the inability to perform full three-way ANOVAs with the EEGLAB study plugin restricted statistical comparisons.

7 Perspectives

Future studies and analyses could address these limitations in several ways.

First, several additional analyses could be conducted using the data already collected in this project, particularly from the training sessions of study 3. In both study 2 and study 3, the Stroop task error rate could be analyzed to provide more qualitative insight into participants' inhibitory control performance.

EEG data also provide a wide range of unexplored possibilities, such as investigating EF electrophysiological activity and its relation to the potential neural correlates of LE.

Further analyses of resting-state, and why not in-task, microstates could be made, using more advance analysis types or using the The ArsQ results (see section 1.5 in Chapter 5).

More precise source localization, examination of tau and mu components, and advanced statistical analyses (e.g., LIMO EEG for multi-factor ANOVAs) could refine understanding of the neural mechanisms already explored in this project. Further analyses regarding ERSPs between TMRs or SNRs could also reveal interaction between alpha dynamics and difficulty in SIN and SIS.

Moreover, exploring correlations between the EEG results and behavioral data would be crucial for the exploration of effortful speech understanding. The relation between subjective LE and alpha dynamics, especially regarding the ICs results, would help clarify how this frequency band is related to speech understanding in effortful conditions. Interaction between alpha *vs.* subjective LE *TMR/SNR* or alpha *vs.* subjective LE *training sessions* would add great insights to this project.

Beyond the current dataset, future experimental protocols could expand the range of SIN and SIS levels (with a larger TMR/SNR range) to provide a more comprehensive understanding of how EFs, especially inhibition, support speech understanding in adverse listening conditions. The use of different noise types, more than two simultaneous talkers, or other corpora could also show complementary results to this project. Cognitive training design could also be refined by using alternative or combined inhibitory control tasks (e.g., go/no-go, stop-signal, or auditory Stroop tasks), which may enhance training efficacy and enable more reliable transfer effects on speech intelligibility and listening effort.

Finally, the analysis pipeline developed during this project provides a foundation that could be reused and improved in future work, supporting both replication and methodological advances in the study of speech understanding in effortful environments.

Conclusion

This manuscript presents a behavioral and electrophysiological investigation of speech listening in effortful situations. Three studies in this project addressed questions about listening in complex auditory environments, the impact of using a native language corpus, its relationship with executive functioning, and the possibility of improving it through cognitive training. A central aim of this project was to investigate EEG alpha brain dynamics during listening in such situations.

Behavioral analyses of speech perception in noise or multi-talker situations allowed us to answer the scientific questions asked at the start of this project. We validated the hypothesis that using a native language positively influences speech intelligibility and listening effort, particularly under adverse conditions. Additionally, we observed a positive correlation between speech understanding and inhibitory control, as measured by the Stroop task, in adverse conditions. And finally, the cognitive training of speech understanding task induced increased intelligibility and decreased listening effort.

Regarding electrophysiological activity, we confirmed that alpha-related processes are involved in listening in effortful situations and that multiple neural generators contribute to this activity.

In conclusion, listening effort constitutes a critical factor in oral communication and should be explicitly considered when investigating speech perception and understanding. By investigating and measuring it through both objective and subjective methods, we can better capture its multidimensional nature. Furthermore, speech research on such topics and their related applications in clinical settings has the potential to improve the understanding and quality of life of individuals with hidden hearing loss or other misdiagnosed auditory impairments.

V

Bibliography

Bibliography

Adair, J. G. (n.d.). The Hawthorne Effect: A Reconsideration of the Methodological Artifact.

Adank, P. (2012). The neural bases of difficult speech comprehension and speech production: Two Activation Likelihood Estimation (ALE) meta-analyses. *Brain and Language*, *122*(1), 42–54. https://doi.org/10.1016/j.bandl.2012.04.014

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology*, *47*(sup2), S53–S71. https://doi.org/10.1080/14992020802301142

Ala, T. S., Alickovic, E., Cabrera, A. F., Whitmer, W. M., Hadley, L. V., Rank, M. L., Lunner, T., & Graversen, C. (2023). Alpha Oscillations During Effortful Continuous Speech: From Scalp EEG to Ear-EEG [Conference Name: IEEE Transactions on Biomedical Engineering]. *IEEE Transactions on Biomedical Engineering*, *70*(4), 1264–1273. https://doi.org/10.1109/TBME.2022.3214428

Alain, C., Du, Y., Bernstein, L. J., Barten, T., & Banai, K. (2018). Listening under difficult conditions: An activation likelihood estimation meta-analysis. *Human Brain Mapping*, *39*(7), 2695–2709. https://doi.org/10.1002/hbm.24031

Alhanbali, S., Dawes, P., Millman, R. E., & Munro, K. J. (2019). Measures of Listening Effort Are Multidimensional. *Ear and Hearing*, *40*(5), 1084–1097. https://doi.org/10.1097/AUD.0000000000000697

Amiri, M., Jarollahi, F., Jalaie, S., & Sameni, S. J. (2020). A New Speech-in-Noise Test for Measuring Informational Masking in Speech Perception Among Elderly Listeners. *Cureus*. https://doi.org/10.7759/cureus.7356

Andeol, G., Guillaume, A., Micheyl, C., Savel, S., Pellieux, L., & Moulin, A. (2011). Auditory Efferents Facilitate Sound Localization in Noise in Humans. *Journal of Neuroscience*, *31*(18), 6759–6763. https://doi.org/10.1523/JNEUROSCI.0248-11.2011

Andéol, G., Suied, C., Scannella, S., & Dehais, F. (2017). The Spatial Release of Cognitive Load in Cocktail Party Is Determined by the Relative Levels of the Talkers. *Journal of the Association for Research in Otolaryngology*, *18*(3), 457–464. https://doi.org/10.1007/s10162-016-0611-7

Arnal, L. H., Poeppel, D., & Giraud, A.-L. (2016). A Neurophysiological Perspective on Speech Processing in “The Neurobiology of Language”. In *Neurobiology of Language* (pp. 463–478). Elsevier. https://doi.org/10.1016/B978-0-12-407794-2.00038-9

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: One decade on. *Trends in Cognitive Sciences*, *18*(4), 177–185. https://doi.org/10.1016/j.tics.2013.12.003

Au, J., Gibson, B. C., Bunarjo, K., Buschkuehl, M., & Jaeggi, S. M. (2020). Quantifying the Difference Between Active and Passive Control Groups in Cognitive Interventions Using Two Meta-analytical Approaches [Company: Springer Distributor: Springer Institution: Springer Label: Springer Publisher: Springer International Publishing]. *Journal of Cognitive Enhancement*, *4*(2), 192–210. https://doi.org/10.1007/s41465-020-00164-6

Baddeley, A. (2000). The episodic buffer: A new component of working memory? [Publisher: Elsevier]. *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. D., & Hitch, G. (1974, January). Working Memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (pp. 47–89, Vol. 8). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

Baggetta, P., & Alexander, P. A. (2016). Conceptualization and Operationalization of Executive Function. *Mind, Brain, and Education*, *10*(1), 10–33. https://doi.org/10.1111/mbe.12100

Bannon, S., Gonsalvez, C. J., Croft, R. J., & Boyce, P. M. (2002). Response inhibition deficits in obsessive-compulsive disorder. *Psychiatry Research*, *110*(2), 165–174. https://doi.org/10.1016/s0165-1781(02)00104-x

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin*, *121*(1), 65–94. https://doi.org/10.1037/0033-2909.121.1.65

Bechara, A. (2005). Decision making, impulse control and loss of willpower to resist drugs: A neurocognitive perspective. *Nature Neuroscience*, *8*(11), 1458–1463. https://doi.org/10.1038/nn1584

Beppi, C., Ribeiro Violante, I., Scott, G., & Sandrone, S. (2021). EEG, MEG and neuromodulatory approaches to explore cognition: Current status and future directions. *Brain and Cognition*, *148*, 105677. https://doi.org/10.1016/j.bandc.2020.105677

Berger, H. (1929). Über das Elektrenkephalogramm des Menschen [Company: Springer Distributor: Springer Institution: Springer Label: Springer Publisher: Springer-Verlag]. *Archiv für Psychiatrie und Nervenkrankheiten*, *87*(1), 527–570. https://doi.org/10.1007/BF01797193

Berkman, E. T., Kahn, L. E., & Merchant, J. S. (2014). Training-Induced Changes in Inhibitory Control Network Activity. *The Journal of Neuroscience*, *34*(1), 149–157. https://doi.org/10.1523/JNEUROSCI.3564-13.2014

Bertoli, S., & Bodmer, D. (2014). Novel sounds as a psychophysiological measure of listening effort in older listeners with and without hearing loss. *Clinical Neurophysiology*, *125*(5), 1030–1041. https://doi.org/10.1016/j.clinph.2013.09.045

Bess, F. H., & Hornsby, B. W. Y. (2014). Commentary: Listening Can Be Exhausting—Fatigue in Children and Adults With Hearing Loss. *Ear and Hearing*, *35*(6), 592. https://doi.org/10.1097/AUD.0000000000000099

Besser, J., Koelewijn, T., Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2013). How Linguistic Closure and Verbal Working Memory Relate to Speech Recognition in Noise—A Review [Publisher: SAGE Publications]. *Trends in Amplification*, *17*(2), 75–93. https://doi.org/10.1177/1084713813495459

Best, V., Ozmeral, E. J., Kopčo, N., & Shinn-Cunningham, B. G. (2008). Object continuity enhances selective auditory attention [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *105*(35), 13174–13178. https://doi.org/10.1073/pnas.0803718105

Billings, C. J., Tremblay, K. L., Stecker, G. C., & Tolin, W. M. (2009). Human evoked cortical activity to signal-to-noise ratio and absolute signal level. *Hearing Research*, *254*(1-2), 15–24. https://doi.org/10.1016/j.heares.2009.04.002

Binder, J. R. (2015). The Wernicke area: Modern evidence and a reinterpretation. *Neurology*, *85*(24), 2170–2175. https://doi.org/10.1212/WNL.0000000000002219

Binder, J. R. (2017). Current Controversies on Wernicke's Area and its Role in Language [Company: Springer Distributor: Springer Institution: Springer Label: Springer Publisher: Springer US]. *Current Neurology and Neuroscience Reports*, *17*(8), 1–10. https://doi.org/10.1007/s11910-017-0764-8

Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience*, *7*(3), 295–301. https://doi.org/10.1038/nn1198

Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *The Journal of the Acoustical Society of America*, *107*(2), 1065–1066. https://doi.org/10.1121/1.428288

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America*, *121*(4), 2339–2349. https://doi.org/10.1121/1.2642103

Brain Mechanisms of Auditory Scene Analysis. (2020, May). In *The Cognitive Neurosciences* (6th ed., pp. 159–166). The MIT Press. https://doi.org/10.7551/mitpress/11442.003.0020

Brännström, K. J., Karlsson, E., Waechter, S., & Kastberg, T. (2018). Listening Effort: Order Effects and Core Executive Functions [Publisher: Thieme Medical Publishers]. *Journal of the American Academy of Audiology*, *29*(8), 734–747. https://doi.org/10.3766/jaaa.17024

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound* (2. paperback ed., repr). MIT Press.

Brehm, J. W., & Self, E. A. (1989). The Intensity of Motivation [Publisher: Annual Reviews]. *Annual Review of Psychology*, *40*(Volume 40, 1989), 109–131. https://doi.org/10.1146/annurev.ps.40.020189.000545

Bronkhorst, A. W. (2015). The cocktail-party problem revisited: Early processing and selection of multi-talker speech. *Attention, Perception, & Psychophysics*, *77*(5), 1465–1487. https://doi.org/10.3758/s13414-015-0882-9

Brouwer, S., Van Engen, K. J., Calandruccio, L., & Bradlow, A. R. (2012). Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content. *The Journal of the Acoustical Society of America*, *131*(2), 1449–1464. https://doi.org/10.1121/1.3675943

Brungart, D. S. (2001a). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America*, *109*(3), 1101–1109. https://doi.org/10.1121/1.1345696

Brungart, D. S. (2001b). Evaluation of speech intelligibility with the coordinate response measure. *The Journal of the Acoustical Society of America*, *109*(5), 2276–2279. https://doi.org/10.1121/1.1357812

Brungart, D. S., & Simpson, B. D. (2007). Cocktail party listening in a dynamic multitalker environment [Company: Springer Distributor: Springer Institution: Springer Label: Springer Number: 1 Publisher: Springer-Verlag]. *Perception & Psychophysics*, *69*(1), 79–91. https://doi.org/10.3758/BF03194455

Burgess, P. W., & Shallice, T. (1996). Response suppression, initiation and strategy use following frontal lobe lesions. *Neuropsychologia*, *34*(4), 263–272. https://doi.org/10.1016/0028-3932(95)00104-2

Buzsáki, G., & Draguhn, A. (2004). Neuronal Oscillations in Cortical Networks. *Science*, *304*(5679), 1926–1929. https://doi.org/10.1126/science.1099745

Carolan, P. J., Heinrich, A., Munro, K. J., & Millman, R. E. (2022). Quantifying the Effects of Motivation on Listening Effort: A Systematic Review and Meta-Analysis. *Trends in Hearing*, *26*, 23312165211059982. https://doi.org/10.1177/23312165211059982

Carretti, B., Cornoldi, C., De Beni, R., & Romanò, M. (2005). Updating in working memory: A comparison of good and poor comprehenders. *Journal of Experimental Child Psychology*, *91*(1), 45–66. https://doi.org/10.1016/j.jecp.2005.01.005

Chan, R. C. K., Shum, D., Toulopoulou, T., & Chen, E. Y. H. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology*, *23*(2), 201–216. https://doi.org/10.1016/j.acn.2007.08.010

Chandrasekaran, B., Tessmer, R., & Gnanateja, G. N. (2022). Subcortical Processing of Speech Sounds [ISSN: 2197-1897]. In *Speech Perception* (pp. 13–44). Springer, Cham. https://doi.org/10.1007/978-3-030-81542-4_2

Chang, E. F., Raygor, K. P., & Berger, M. S. (2015). Contemporary model of language organization: An overview for neurosurgeons. *Journal of Neurosurgery*, *122*(2), 250–261. https://doi.org/10.3171/2014.10.JNS132647

Chavan, C. F., Mouthon, M., Draganski, B., van der Zwaag, W., & Spierer, L. (2015). Differential patterns of functional and structural plasticity within and between inferior frontal gyri support training-induced improvements in inhibitory control proficiency. *Human Brain Mapping*, *36*(7), 2527–2543. https://doi.org/10.1002/hbm.22789

Chenot, Q., Hamery, C., Truninger, M., Langer, N., De boissezon, X., & Scannella, S. (2024). Investigating the relationship between resting-state EEG microstates and executive functions: A null finding. *Cortex*. https://doi.org/10.1016/j.cortex.2024.05.019

Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America*, *25*(5), 975–979. https://doi.org/10.1121/1.1907229

Choi, I., Rajaram, S., Varghese, L. A., & Shinn-Cunningham, B. G. (2013). Quantifying attentional modulation of auditory-evoked cortical responses from single-trial electroencephalography. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00115

Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America*, *123*(1), 414–427. https://doi.org/10.1121/1.2804952

Cooke, M., & Lecumberri, M. L. G. (2012). The intelligibility of Lombard speech for non-native listeners. *The Journal of the Acoustical Society of America*, *132*(2), 1120–1129. https://doi.org/10.1121/1.4732062

Cristofori, I., Cohen-Zimerman, S., & Grafman, J. (2019, January). Chapter 11 - Executive functions. In M. D'Esposito & J. H. Grafman (Eds.), *Handbook of Clinical Neurology* (pp. 197–219, Vol. 163). Elsevier. https://doi.org/10.1016/B978-0-12-804281-6.00011-2

Culling, J. F., & Stone, M. A. (2017). Energetic Masking and Masking Release [ISSN: 2197-1897]. In *The Auditory System at the Cocktail Party* (pp. 41–73). Springer, Cham. https://doi.org/10.1007/978-3-319-51662-2_3

Curio, G., Mackert, B.-M., Burghoff, M., Koetitz, R., Abraham-Fuchs, K., & Härer, W. (1994). Localization of evoked neuromagnetic 600 Hz activity in the cerebral somatosensory system. *Electroencephalography and Clinical Neurophysiology*, *91*(6), 483–487. https://doi.org/10.1016/0013-4694(94)90169-4

Damasio, H., Grabowski, T., Frank, R., Galaburda, A. M., & Damasio, A. R. (1994). The return of Phineas Gage: Clues about the brain from the skull of a famous patient. *Science (New York, N.Y.)*, *264*(5162), 1102–1105. https://doi.org/10.1126/science.8178168

Darwin, C. J. (2008). Spatial Hearing and Perceiving Sources. In *Auditory Perception of Sound Sources* (pp. 215–232). Springer, Boston, MA. https://doi.org/10.1007/978-0-387-71305-2_8

Davidson, D. J., Zacks, R. T., & Williams, C. C. (2003). Stroop Interference, Practice, and Aging. *Neuropsychology, development, and cognition. Section B, Aging, neuropsychology and cognition*, *10*(2), 85–98. https://doi.org/10.1076/anec.10.2.85.14463

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research*, *229*(1), 132–147. https://doi.org/10.1016/j.heares.2007.01.014

Deco, G., Jirsa, V. K., & McIntosh, A. R. (2011). Emerging concepts for the dynamical organization of resting-state activity in the brain [Publisher: Nature Publishing Group]. *Nature Reviews Neuroscience*, *12*(1), 43–56. https://doi.org/10.1038/nrn2961

Dehais, F., Duprès, A., Blum, S., Drougard, N., Scannella, S., Roy, R. N., & Lotte, F. (2019). Monitoring Pilot's Mental Workload Using ERPs and Spectral Power with a Six-Dry-Electrode EEG System in Real Flight Conditions [Publisher: Multidisciplinary Digital Publishing Institute]. *Sensors*, *19*(6), 1324. https://doi.org/10.3390/s19061324

Dehais, F., Lafont, A., Roy, R., & Fairclough, S. (2020). A Neuroergonomics Approach to Mental Workload, Engagement and Human Performance [Publisher: Frontiers]. *Frontiers in Neuroscience*, *14*. https://doi.org/10.3389/fnins.2020.00268

Delorme, A. (2023). EEG is better left alone [Publisher: Nature Publishing Group]. *Scientific Reports*, *13*(1), 2372. https://doi.org/10.1038/s41598-023-27528-0

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Denes, P. B., & Pinson, E. N. (1993). The Speech Chain. In *The Speech Chain: The Physics and Biology of Spoken Language*.

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Diaz, B. A., Van Der Sluis, S., Moens, S., Benjamins, J. S., Migliorati, F., Stoffers, D., Den Braber, A., Poil, S.-S., Hardstone, R., Van 't Ent, D., Boomsma, D. I., De Geus, E., Mansvelder, H. D., Van Someren, E. J. W., & Linkenkaer-Hansen, K. (2013). The Amsterdam Resting-State Questionnaire reveals multiple phenotypes of resting-state cognition [Publisher: Frontiers]. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00446

Dimitrijevic, A., Smith, M. L., Kadis, D. S., & Moore, D. R. (2017). Cortical Alpha Oscillations Predict Speech Intelligibility. *Frontiers in Human Neuroscience*, *11*. Retrieved October 24, 2023, from https://www.frontiersin.org/articles/10.3389/fnhum.2017.00088

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

Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience*, *8*. Retrieved June 19, 2023, from https://www.frontiersin.org/articles/10.3389/fnhum.2014.00311

DiNino, M., Holt, L. L., & Shinn-Cunningham, B. G. (2022). Cutting Through the Noise: Noise-Induced Cochlear Synaptopathy and Individual Differences in Speech Understanding Among Listeners With Normal Audiograms. *Ear and Hearing*, *43*(1), 9. https://doi.org/10.1097/AUD.0000000000001147

Dotson, V. M., Sozda, C. N., Marsiske, M., & Perlstein, W. M. (2013). Within-session Practice Eliminates Age Differences in Cognitive Control. *Neuropsychology, development, and cognition. Section B, Aging, neuropsychology and cognition*, *20*(5), 522–531. https://doi.org/10.1080/13825585.2012.736469

Duffy, J. D., & Campbell III, J. J. (2001). Regional prefrontal syndromes: A theoretical and clinical overview. In *The frontal lobes and neuropsychiatric illness* (pp. 113–123). American Psychiatric Publishing, Inc.

Duke, L. M., & Kaszniak, A. W. (2000). Executive Control Functions in Degenerative Dementias: A Comparative Review [Company: Springer Distributor: Springer Institution: Springer Label: Springer Publisher: Kluwer Academic Publishers-Plenum Publishers]. *Neuropsychology Review*, *10*(2), 75–99. https://doi.org/10.1023/A:1009096603879

Duncan, J., Assem, M., & Shashidhara!, S. (2020). Integrated intelligence from distributed brain activity. *Trends in cognitive sciences*, *24*(10), 838–852. https://doi.org/10.1016/j.tics.2020.06.012

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences*, *23*(10), 475–483. https://doi.org/10.1016/S0166-2236(00)01633-7

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., & Kidd, G. (2003). Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *The Journal of the Acoustical Society of America*, *114*(1), 368–379. https://doi.org/10.1121/1.1577562

Eckert, M. A., Teubner-Rhodes, S., & Vaden, K. I. J. (2016). Is Listening in Noise Worth It? The Neurobiology of Speech Recognition in Challenging Listening Conditions. *Ear and Hearing*, *37*, 101S. https://doi.org/10.1097/AUD.0000000000000300

Eisner, F., McGettigan, C., Faulkner, A., Rosen, S., & Scott, S. K. (2010). Inferior Frontal Gyrus Activation Predicts Individual Differences in Perceptual Learning of Cochlear-Implant Simulations [Publisher: Society for Neuroscience Section: Articles]. *Journal of Neuroscience*, *30*(21), 7179–7186. https://doi.org/10.1523/JNEUROSCI.4040-09.2010

Ellis, R. J., & Rönnberg, J. (2014). Cognition and Speech-In-Noise Recognition: The Role of Proactive Interference. *Journal of the American Academy of Audiology*, *25*(10), 975–982. https://doi.org/10.3766/jaaa.25.10.6

Enge, S., Behnke, A., Fleischhauer, M., Küttler, L., Kliegel, M., & Strobel, A. (2014). No evidence for true training and transfer effects after inhibitory control training in young healthy adults [Place: US Publisher: American Psychological Association]. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 987–1001. https://doi.org/10.1037/a0036165

Eqlimi, E., Bockstael, A., Schönwiesner, M., Talsma, D., & Botteldooren, D. (2023). Time course of EEG complexity reflects attentional engagement during listening to speech in noise. *European Journal of Neuroscience*, *58*(9), 4043–4069. https://doi.org/10.1111/ejn.16159

Ershaid, H., Lizarazu, M., McLaughlin, D., Cooke, M., Simantiraki, O., Koutsogiannaki, M., & Lallier, M. (2024). Contributions of listening effort and intelligibility to cortical tracking of speech in adverse listening conditions. *Cortex*, *172*, 54–71. https://doi.org/10.1016/j.cortex.2023.11.018

Evans, S., McGettigan, C., Agnew, Z., Rosen, S., & Scott, S. (2016). Getting the cocktail party started: Masking effects in speech perception. *Journal of cognitive neuroscience*, *28*(3), 483–500. https://doi.org/10.1162/jocn_a_00913

Fedorenko, E. (2014). The role of domain-general cognitive control in language comprehension. *Frontiers in Psychology*, *5*, 335. https://doi.org/10.3389/fpsyg.2014.00335

Fedorenko, E., Duncan, J., & Kanwisher, N. (2012). Language Selective and Domain-General Regions Lie Side by Side within Broca's Area. *Current Biology*, *22*(21), 2059–2062. https://doi.org/10.1016/j.cub.2012.09.011

Fedorenko, E., & Shain, C. (2021). Similarity of computations across domains does not imply shared implementation: The case of language comprehension. *Current directions in psychological science*, *30*(6), 526–534. https://doi.org/10.1177/09637214211046955

Foroughi, C. K., Monfort, S. S., Paczynski, M., McKnight, P. E., & Greenwood, P. M. (2016). Placebo effects in cognitive training. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(27), 7470–7474. https://doi.org/10.1073/pnas.1601243113

Francis, A. L., Bent, T., Schumaker, J., Love, J., & Silbert, N. (2021). Listener characteristics differentially affect self-reported and physiological measures of effort associated with two challenging listening conditions [Company: Springer Distributor: Springer Institution: Springer Label: Springer Number: 4 Publisher: Springer US]. *Attention, Perception, & Psychophysics*, *83*(4), 1818–1841. https://doi.org/10.3758/s13414-020-02195-9

Francis, A. L., & Love, J. (2020). Listening effort: Are we measuring cognition or affect, or both? *WIREs Cognitive Science*, *11*(1). https://doi.org/10.1002/wcs.1514

Friederici, A. D. (2011). The Brain Basis of Language Processing: From Structure to Function [Publisher: American Physiological Society]. *Physiological Reviews*, *91*(4), 1357–1392. https://doi.org/10.1152/physrev.00006.2011

Friederici, A. D., & Gierhan, S. M. (2013). The language network. *Current Opinion in Neurobiology*, *23*(2), 250–254. https://doi.org/10.1016/j.conb.2012.10.002

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, *137*(2), 201–225. https://doi.org/10.1037/0096-3445.137.2.201

Füllgrabe, C., & Rosen, S. (2016). On The (Un)importance of Working Memory in Speech-in-Noise Processing for Listeners with Normal Hearing Thresholds [Publisher: Frontiers]. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01268

Gagné, J.-P., Besser, J., & Lemke, U. (2017). Behavioral Assessment of Listening Effort Using a Dual-Task Paradigm. *Trends in Hearing*, *21*, 2331216516687287. https://doi.org/10.1177/2331216516687287

Gatehouse, S., & Noble, W. (2004). The Speech, Spatial and Qualities of Hearing Scale (SSQ). *International Journal of Audiology*, *43*(2), 85–99. https://doi.org/10.1080/14992020400050014

Gathercole, S. E., Dunning, D. L., Holmes, J., & Norris, D. (2016). Working memory training involves learning new skills. *Journal of memory and language*, *105*, 19–42. https://doi.org/10.1016/j.jml.2018.10.003

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations [Number: 4 Publisher: Nature Publishing Group]. *Nature Neuroscience*, *15*(4), 511–517. https://doi.org/10.1038/nn.3063

Gobet, F., & Sala, G. (2023). Cognitive Training: A Field in Search of a Phenomenon [Publisher: SAGE Publications Inc]. *Perspectives on Psychological Science*, *18*(1), 125–141. https://doi.org/10.1177/17456916221091830

Golestani, N., Rosen, S., & Scott, S. K. (2009). Native-language benefit for understanding speech-in-noise: The contribution of semantics*. *Bilingualism: Language and Cognition*, *12*(03), 385. https://doi.org/10.1017/S1366728909990150

Grech, R., Cassar, T., Muscat, J., Camilleri, K. P., Fabri, S. G., Zervakis, M., Xanthopoulos, P., Sakkalis, V., & Vanrumste, B. (2008). Review on solving the inverse problem in EEG source analysis [Publisher: BioMed Central]. *Journal of NeuroEngineering and Rehabilitation*, *5*(1), 1–33. https://doi.org/10.1186/1743-0003-5-25

Hall, A. J., Winneke, A., & Rennies-Hochmuth, J. (2019). EEG alpha power as a measure of listening effort reduction in adverse conditions.

Hällgren, M., Larsby, B., & Arlinger, S. (2006). A Swedish version of the Hearing In Noise Test (HINT) for measurement of speech recognition. *International Journal of Audiology*, *45*(4), 227–237. https://doi.org/10.1080/14992020500429583

Hambrook, D. A., & Tata, M. S. (2019). The effects of distractor set-size on neural tracking of attended speech. *Brain and Language*, *190*, 1–9. https://doi.org/10.1016/j.bandl.2018.12.005

Hamery, C., Andeol, G., Scannella, S., & Isnard, V. (2023, December). Influence of the language proficiency on speech intelligibility and listening effort in multi-talker situations. https://doi.org/10.13140/RG.2.2.16597.44000

Hari, R., & Puce, A. (2023). *MEG-EEG Primer* [Google-Books-ID: wRrOEAAAQBAJ]. Oxford University Press.

Harlow, J. M. (1869). *Recovery from the passage of an iron bar through the head.*

Harris, D. J., Wilson, M. R., Chillingsworth, K., Mitchell, G., Smith, S., Arthur, T., Brock, K., & Vine, S. J. (2023). Can cognitive training capitalise on near transfer effects? Limited evidence of transfer following online inhibition training in a randomised-controlled trial (C. Andreu-Sánchez, Ed.). *PLOS ONE*, *18*(11), e0293657. https://doi.org/10.1371/journal.pone.0293657

Hart, S. G., & Staveland, L. E. (1988, January). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology* (pp. 139–183, Vol. 52). North-Holland. https://doi.org/10.1016/S0166-4115(08)62386-9

Hartmann, L., Sallard, E., & Spierer, L. (2016). Enhancing frontal top-down inhibitory control with Go/NoGo training. *Brain Structure and Function*, *221*(7), 3835–3842. https://doi.org/10.1007/s00429-015-1131-7

He, B., & Lian, J. (2005). Electrophysiological Neuroimaging. In *Neural Engineering* (pp. 221–261). Springer, Boston, MA. https://doi.org/10.1007/0-306-48610-5_7

He, B., Sohrabpour, A., Brown, E., & Liu, Z. (2018). Electrophysiological Source Imaging: A Noninvasive Window to Brain Dynamics [Publisher: Annual Reviews]. *Annual Review of Biomedical Engineering*, *20*(Volume 20, 2018), 171–196. https://doi.org/10.1146/annurev-bioeng-062117-120853

Herrmann, B., & Johnsrude, I. S. (2020). A model of listening engagement (MoLE). *Hearing Research*, *397*, 108016. https://doi.org/10.1016/j.heares.2020.108016

Herrmann, C. S., & Demiralp, T. (2005). Human EEG gamma oscillations in neuropsychiatric disorders. *Clinical Neurophysiology*, *116*(12), 2719–2733. https://doi.org/10.1016/j.clinph.2005.07.007

Herrmann, C. S., Munk, M. H. J., & Engel, A. K. (2004). Cognitive functions of gamma-band activity: Memory match and utilization [Publisher: Elsevier]. *Trends in Cognitive Sciences*, *8*(8), 347–355. https://doi.org/10.1016/j.tics.2004.06.006

Hertrich, I., Dietrich, S., & Ackermann, H. (2020). The Margins of the Language Network in the Brain [Publisher: Frontiers]. *Frontiers in Communication*, *5*. https://doi.org/10.3389/fcomm.2020.519955

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*(1), 67–99. https://doi.org/10.1016/j.cognition.2003.10.011

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. https://doi.org/10.1038/nrn2113

Hogervorst, M. A., Brouwer, A.-M., & van Erp, J. B. F. (2014). Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. *Frontiers in Neuroscience*, *8*, 322. https://doi.org/10.3389/fnins.2014.00322

Holman, J. A., Drummond, A., & Naylor, G. (2021). The Effect of Hearing Loss and Hearing Device Fitting on Fatigue in Adults: A Systematic Review. *Ear and Hearing*, *42*(1), 1. https://doi.org/10.1097/AUD.0000000000000909

Holt, L. L., Peelle, J. E., Coffin, A. B., Popper, A. N., & Fay, R. R. (Eds.). (2022). *Speech Perception* (Vol. 74). Springer International Publishing. Retrieved May 6, 2024, from DOI:%2010.1007/978-3-030-81542-4

Hopstaken, J. F., Van Der Linden, D., Bakker, A. B., & Kompier, M. A. J. (2015). A multifaceted investigation of the link between mental fatigue and task disengagement. *Psychophysiology*, *52*(3), 305–315. https://doi.org/10.1111/psyp.12339

Horton, C., D'Zmura, M., & Srinivasan, R. (2013). Suppression of competing speech through entrainment of cortical oscillations. *Journal of Neurophysiology*, *109*(12), 3082–3093. https://doi.org/10.1152/jn.01026.2012

Horton, C., Srinivasan, R., & D'Zmura, M. (2014). Envelope responses in single-trial EEG indicate attended speaker in a 'cocktail party'. *Journal of Neural Engineering*, *11*(4), 046015. https://doi.org/10.1088/1741-2560/11/4/046015

Houben, K. (2011). Overcoming the urge to splurge: Influencing eating behavior by manipulating inhibitory control. *Journal of Behavior Therapy and Experimental Psychiatry*, *42*(3), 384–388. https://doi.org/10.1016/j.jbtep.2011.02.008

Houben, K., Havermans, R. C., Nederkoorn, C., & Jansen, A. (2012). Beer à no-go: Learning to stop responding to alcohol cues reduces alcohol intake via reduced affective associations rather than increased response inhibition. *Addiction*, *107*(7), 1280–1287. https://doi.org/10.1111/j.1360-0443.2012.03827.x

Houben, K., Nederkoorn, C., Wiers, R. W., & Jansen, A. (2011). Resisting temptation: Decreasing alcohol-related affect and drinking behavior by training response inhibition. *Drug and Alcohol Dependence*, *116*(1), 132–136. https://doi.org/10.1016/j.drugalcdep.2010.12.011

House, A. S., Williams, C., Hecker, M. H. L., & Kryter, K. D. (1963). Psychoacoustic Speech Tests: A Modified Rhyme Test. *The Journal of the Acoustical Society of America*, *35*, 1899. https://doi.org/10.1121/1.2142744

Hunter, C. R. (2020). Tracking Cognitive Spare Capacity During Speech Perception With EEG/ERP: Effects of Cognitive Load and Sentence Predictability. *Ear & Hearing*, *41*(5), 1144–1157. https://doi.org/10.1097/AUD.0000000000000856

Ingvalson, E. M., Dhar, S., Wong, P. C. M., & Liu, H. (2015). Working memory training to improve speech perception in noise across languages. *The Journal of the Acoustical Society of America*, *137*(6), 3477–3486. https://doi.org/10.1121/1.4921601

Ingvalson, E. M., Lansford, K. L., Fedorova, V., & Fernandez, G. (2017). Receptive Vocabulary, Cognitive Flexibility, and Inhibitory Control Differentially Predict Older and Younger Adults' Success Perceiving Speech by Talkers With Dysarthria [Publisher: American Speech-Language-Hearing Association]. *Journal of Speech, Language, and Hearing Research*, *60*(12), 3632–3641. https://doi.org/10.1044/2017_JSLHR-H-17-0119

Ishii, R., Shinosaki, K., Ukai, S., Inouye, T., Ishihara, T., Yoshimine, T., Hirabuki, N., Asada, H., Kihara, T., Robinson, S. E., & Takeda, M. (1999). Medial prefrontal cortex generates frontal midline theta rhythm. *Neuroreport*, *10*(4), 675–679. https://doi.org/10.1097/00001756-199903170-00003

Isnard, V., Chastres, V., & Andéol, G. (2024). French version of the coordinate response measure corpus and its validation on a speech-on-speech task. *JASA Express Letters*, *4*(7), 075203. https://doi.org/10.1121/10.0028059

Jaeggi, S. M., Karbach, J., & Strobach, T. (2017). Editorial Special Topic: Enhancing Brain and Cognition Through Cognitive Training [Company: Springer Distributor: Springer Institution: Springer Label: Springer Publisher: Springer International Publishing]. *Journal of Cognitive Enhancement*, *1*(4), 353–357. https://doi.org/10.1007/s41465-017-0057-9

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning — implications for training and transfer. *Intelligence*, *38*(6), 625–635. https://doi.org/10.1016/j.intell.2010.09.001

Jasper, H. H., & Andrew, H. L. (1938). BRAIN POTENTIALS AND VOLUNTARY MUSCLE ACTIVITY IN MAN — Journal of Neurophysiology — American Physiological Society. Retrieved August 21, 2025, from https://journals.physiology.org/doi/abs/10.1152/jn.1938.1.2.87

Jensen, O., & Mazaheri, A. (2010). Shaping Functional Architecture by Oscillatory Alpha Activity: Gating by Inhibition [Publisher: Frontiers]. *Frontiers in Human Neuroscience*, *4*. https://doi.org/10.3389/fnhum.2010.00186

Jenson, D., Bowers, A. L., Harkrider, A. W., Thornton, D., Cuellar, M., & Saltuklaroglu, T. (2014). Temporal dynamics of sensorimotor integration in speech perception and production: Independent component analysis of EEG data [Publisher: Frontiers]. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00656

Jenson, D., Harkrider, A. W., Thornton, D., Bowers, A. L., & Saltuklaroglu, T. (2015). Auditory cortical deactivation during speech production and following speech perception: An EEG investigation of the temporal dynamics of the auditory alpha rhythm. *Frontiers in Human Neuroscience*, *9*, 534. https://doi.org/10.3389/fnhum.2015.00534

Jenson, D., Thornton, D., Saltuklaroglu, T., & Harkrider, A. (2014). Speech perception, production, and the sensorimotor mu rhythm. *Proceedings of the 2014 Biomedical Sciences and Engineering Conference*, 1–4. https://doi.org/10.1109/BSEC.2014.6867736

Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., & Carlyon, R. P. (2013). Swinging at a Cocktail Party: Voice Familiarity Aids Speech Perception in the Presence of a Competing Voice [Publisher: SAGE Publications Inc]. *Psychological Science*, *24*(10), 1995–2004. https://doi.org/10.1177/0956797613482467

Johnsrude, I. S., & Rodd, J. M. (2016, January). Chapter 40 - Factors That Increase Processing Demands When Listening to Speech. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 491–502). Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00040-7

Kandell, E. R., Schwartz, J. H., Jessel, T. M., Siegelbaum, S. A., & Hudspeth, A. J. (2000). *Principles of Neural Sciences* (Fifth Edition).

Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Developmental Science*, *12*(6), 978–990. https://doi.org/10.1111/j.1467-7687.2009.00846.x

Karbach, J., & Kray, J. (2021). Executive Function Training. In *Cognitive Training* (pp. 199–212). Springer, Cham. https://doi.org/10.1007/978-3-030-39292-5_14

Katz, B., Jones, M. R., Shah, P., Buschkuehl, M., & Jaeggi, S. M. (2021). Individual Differences in Cognitive Training Research. In *Cognitive Training* (pp. 107–123). Springer, Cham. https://doi.org/10.1007/978-3-030-39292-5_8

Keur-Huizinga, L., Kramer, S. E., de Geus, E. J. C., & Zekveld, A. A. (2024). A Multimodal Approach to Measuring Listening Effort: A Systematic Review on the Effects of Auditory Task Demand on Physiological Measures and Their Relationship. *Ear and Hearing*, *45*(5), 1089. https://doi.org/10.1097/AUD.0000000000001508

Kidd, G., & Colburn, H. S. (2017). Informational Masking in Speech Recognition [ISSN: 2197-1897]. In *The Auditory System at the Cocktail Party* (pp. 75–109). Springer, Cham. https://doi.org/10.1007/978-3-319-51662-2_4

Kitterick, P. T., Bailey, P. J., & Summerfield, A. Q. (2010). Benefits of knowing who, where, and when in multi-talker listening. *The Journal of the Acoustical Society of America*, *127*(4), 2498–2508. https://doi.org/10.1121/1.3327507

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: The inhibition–timing hypothesis. *Brain Research Reviews*, *53*(1), 63–88. https://doi.org/10.1016/j.brainresrev.2006.06.003

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., Gillberg, C. G., Forssberg, H., & Westerberg, H. (2005). Computerized Training of Working Memory in Children With ADHD-A Randomized, Controlled Trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, *44*(2), 177–186. https://doi.org/10.1097/00004583-200502000-00010

Knight, S., & Heinrich, A. (2017). Different Measures of Auditory and Visual Stroop Interference and Their Relationship to Speech Intelligibility in Noise. *Frontiers in Psychology*, *8*. Retrieved October 11, 2022, from https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00230

Knight, S., & Heinrich, A. (2019). Visual Inhibition Measures Predict Speech-in-Noise Perception Only in People With Low Levels of Education. *Frontiers in Psychology*, *9*. Retrieved October 11, 2022, from https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02779

Knyazev, G. G. (2012). EEG delta oscillations as a correlate of basic homeostatic and motivational processes. *Neuroscience & Biobehavioral Reviews*, *36*(1), 677–695. https://doi.org/10.1016/j.neubiorev.2011.10.002

Koch, I., Lawo, V., Fels, J., & Vorländer, M. (2011). Switching in the cocktail party: Exploring intentional control of auditory selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(4), 1140–1147. https://doi.org/10.1037/a0022189

Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2012). Pupil Dilation Uncovers Extra Listening Effort in the Presence of a Single-Talker Masker. *Ear and Hearing*, *33*(2), 291. https://doi.org/10.1097/AUD.0b013e3182310019

Kollmeier, B., Warzybok, A., Hochmuth, S., Zokoll, M. A., Uslar, V., Brand, T., & Wagener, K. C. (2015). The multilingual matrix test: Principles, applications, and comparison across languages: A review [Publisher: Informa UK Limited]. *International Journal of Audiology*, *54*(sup2), 3–16. https://doi.org/10.3109/14992027.2015.1020971

Kramer, S. E., Teunissen, C. E., & Zekveld, A. A. (2016). Cortisol, Chromogranin A, and Pupillary Responses Evoked by Speech Recognition Tasks in Normally Hearing and Hard-of-Hearing Listeners: A Pilot Study. *Ear and Hearing*, *37*, 126S. https://doi.org/10.1097/AUD.0000000000000311

Krueger, M., Schulte, M., Brand, T., & Holube, I. (2017). Development of an adaptive scaling method for subjective listening effort. *The Journal of the Acoustical Society of America*, *141*(6), 4680–4693. https://doi.org/10.1121/1.4986938

Krueger, M., Schulte, M., Zokoll, M. A., Wagener, K. C., Meis, M., Brand, T., & Holube, I. (2017). Relation Between Listening Effort and Speech Intelligibility in Noise. *American Journal of Audiology*, *26*(3S), 378–392. https://doi.org/10.1044/2017_AJA-16-0136

Lanzilotti, C. (2021, November). *Étude multidimensionnelle de l'effort d'écoute en situations multilocuteurs : Mesures subjectives, comportementales et en imagerie optique* [Doctoral dissertation, Sorbonne Université]. Retrieved October 10, 2022, from https://tel.archives-ouvertes.fr/tel-03533020

Lanzilotti, C., Andéol, G., Micheyl, C., & Scannella, S. (2022). Cocktail party training induces increased speech intelligibility and decreased cortical activity in bilateral inferior frontal gyri. A functional near-infrared study. *PLOS ONE*, *17*(12), e0277801. https://doi.org/10.1371/journal.pone.0277801

Lau, M. K., Hicks, C., Kroll, T., & Zupancic, S. (2019). Effect of Auditory Task Type on Physiological and Subjective Measures of Listening Effort in Individuals With Normal Hearing [Publisher: American Speech-Language-Hearing Association]. *Journal of Speech, Language, and Hearing Research*, *62*(5), 1549–1560. https://doi.org/10.1044/2018_JSLHR-H-17-0473

Lawrence, R. J., Wiggins, I. M., Anderson, C. A., Davies-Thompson, J., & Hartley, D. E. H. (2018). Cortical correlates of speech intelligibility measured using functional near-infrared spectroscopy (fNIRS). *Hearing Research*, *370*, 53–64. https://doi.org/10.1016/j.heares.2018.09.005

Lebely, C., Lepron, E., Bigarre, I., Hamery, C., De Boissezon, X., & Scannella, S. (2024). EEG Spectral Power Changes in Patients With Dysexecutive Syndrome Following Cognitive Intervention. *Brain and Behavior*, *14*(11), e70148. https://doi.org/10.1002/brb3.70148

Lecumberri, M. L. G., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *The Journal of the Acoustical Society of America*, *119*(4), 2445–2454. https://doi.org/10.1121/1.2180210

Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*, *52*(11), 864–886. https://doi.org/10.1016/j.specom.2010.08.014

Lehmann, D., Ozaki, H., & Pal, I. (1987). EEG alpha map series: Brain micro-states by space-oriented adaptive segmentation. *Electroencephalography and Clinical Neurophysiology*, *67*(3), 271–288. https://doi.org/10.1016/0013-4694(87)90025-3

Lehtelä, L., Salmelin, R., & Hari, R. (1997). Evidence for reactive magnetic 10-Hz rhythm in the human auditory cortex. *Neuroscience Letters*, *222*(2), 111–114. https://doi.org/10.1016/S0304-3940(97)13361-4

Lelo de Larrea-Mancera, E. S., Solís-Vivanco, R., Sánchez-Jimenez, Y., Coco, L., Gallun, F. J., & Seitz, A. R. (2023). Development and validation of a Spanish-language spatial release from masking task in a Mexican population [Publisher: Acoustical Society of America]. *The Journal of the Acoustical Society of America*, *153*(1), 316–327. https://doi.org/10.1121/10.0016850

Lenartowicz, A., Verbruggen, F., Logan, G. D., & Poldrack, R. A. (2011). Inhibition-related Activation in the Right Inferior Frontal Gyrus in the Absence of Inhibitory Cues. *Journal of Cognitive Neuroscience*, *23*(11), 3388–3399. https://doi.org/10.1162/jocn_a_00031

Leske, S., Tse, A., Oosterhof, N. N., Hartmann, T., Müller, N., Keil, J., & Weisz, N. (2014). The strength of alpha and beta oscillations parametrically scale with the strength of an illusory auditory percept. *NeuroImage*, *88*, 69–78. https://doi.org/10.1016/j.neuroimage.2013.11.014

Lewald, J., & Getzmann, S. (2015). Electrophysiological correlates of cocktail-party listening. *Behavioural Brain Research*, *292*, 157–166. https://doi.org/10.1016/j.bbr.2015.06.025

Lezak, M. D. (1982). The Problem of Assessing Executive Functions. *International Journal of Psychology*, *17*(1-4), 281–297. https://doi.org/10.1080/00207598208247445

Liberman, M. C., Epstein, M. J., Cleveland, S. S., Wang, H., & Maison, S. F. (2016). Toward a Differential Diagnosis of Hidden Hearing Loss in Humans [Publisher: Public Library of Science]. *PLOS ONE*, *11*(9), e0162726. https://doi.org/10.1371/journal.pone.0162726

Liberman, M. C., & Kujawa, S. G. (2017). Cochlear synaptopathy in acquired sensorineural hearing loss: Manifestations and mechanisms. *Hearing Research*, *349*, 138–147. https://doi.org/10.1016/j.heares.2017.01.003

Lin, G., & Carlile, S. (2015). Costs of switching auditory spatial attention in following conversational turn-taking [Publisher: Frontiers]. *Frontiers in Neuroscience*, *9*. https://doi.org/10.3389/fnins.2015.00124

Liu, C., Zhou, C., Wang, J., Fietkiewicz, C., & Loparo, K. A. (2020). The role of coupling connections in a model of the cortico-basal ganglia-thalamocortical neural loop for the generation of beta oscillations. *Neural Networks*, *123*, 381–392. https://doi.org/10.1016/j.neunet.2019.12.021

Lu, L., Ren, Y., Yu, T., Liu, Z., Wang, S., Tan, L., Zeng, J., Feng, Q., Lin, R., Liu, Y., Guo, Q., & Luo, M. (2020). Control of locomotor speed, arousal, and hippocampal theta rhythms by the nucleus incertus [Publisher: Nature Publishing Group]. *Nature Communications*, *11*(1), 262. https://doi.org/10.1038/s41467-019-14116-y

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique, second edition*. MIT Press.

Lunner, T. (2003). Cognitive function in relation to hearing aid use. *International Journal of Audiology*, *42*(sup1), 49–58. https://doi.org/10.3109/14992020309074624

Luria, A. R. (1966). *Higher Cortical Functions in Man*. Springer US. https://doi.org/10.1007/978-1-4684-7741-2

Luts, H., Eneman, K., Wouters, J., Schulte, M., Vormann, M., Buechler, M., Dillier, N., Houben, R., Dreschler, W. A., Froehlich, M., Puder, H., Grimm, G., Hohmann, V., Leijon, A., Lombard, A., Mauler, D., & Spriet, A. (2010). Multicenter evaluation of signal enhancement algorithms for hearing aidsa). *The Journal of the Acoustical Society of America*, *127*(3), 1491–1505. https://doi.org/10.1121/1.3299168

MacGregor, L. J., Gilbert, R. A., Balewski, Z., Mitchell, D. J., Erzinçlioğlu, S. W., Rodd, J. M., Duncan, J., Fedorenko, E., & Davis, M. H. (2022). Causal Contributions of the Domain-General (Multiple Demand) and the Language-Selective

Brain Networks to Perceptual and Semantic Challenges in Speech Comprehension. *Neurobiology of Language*, *3*(4), 665–698. https://doi.org/10.1162/nol_a_00081

Mackersie, C. L., & Cones, H. (2011). Subjective and psychophysiological indices of listening effort in a competing-talker task. *Journal of the American Academy of Audiology*, *22*(2), 113–122. https://doi.org/10.3766/jaaa.22.2.6

Mackersie, C. L., MacPhee, I. X., & Heldt, E. W. (2015). Effects of Hearing Loss on Heart-Rate Variability and Skin Conductance Measured During Sentence Recognition in Noise. *Ear and hearing*, *36*(1), 145–154. https://doi.org/10.1097/AUD.0000000000000091

Makeig, S. (1993). Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalography and Clinical Neurophysiology*, *86*(4), 283–293. https://doi.org/10.1016/0013-4694(93)90110-H

Makeig, S., Bell, A. J., Jung, T.-P., & Sejnowski, T. J. (1995). Independent Component Analysis of Electroencephalographic Data. *Advances in Neural Information Processing Systems*.

Makeig, S., Debener, S., Onton, J., & Delorme, A. (2004). Mining event-related brain dynamics [Publisher: Elsevier]. *Trends in Cognitive Sciences*, *8*(5), 204–210. https://doi.org/10.1016/j.tics.2004.03.008

Manuel, A. L., Bernasconi, F., & Spierer, L. (2013). Plastic modifications within inhibitory control networks induced by practicing a stop-signal task: An electrical neuroimaging study. *Cortex*, *49*(4), 1141–1147. https://doi.org/10.1016/j.cortex.2012.12.009

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Matthen, M. (2016). Effort and Displeasure in People Who Are Hard of Hearing. *Ear & Hearing*, *37*(1), 28S–34S. https://doi.org/10.1097/AUD.0000000000000292

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*(7-8), 953–978. https://doi.org/10.1080/01690965.2012.705006

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of Multiple Speech Segmentation Cues: A Hierarchical Framework [Place: US Publisher: American Psychological Association]. *Journal of Experimental Psychology: General*, *134*(4), 477–500. https://doi.org/10.1037/0096-3445.134.4.477

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. *International Journal of Audiology*, *53*(7), 433–445. https://doi.org/10.3109/14992027.2014.890296

McMahon, C. M., Boisvert, I., de Lissa, P., Granger, L., Ibrahim, R., Lo, C. Y., Miles, K., & Graham, P. L. (2016). Monitoring Alpha Oscillations and Pupil Dilation across a Performance-Intensity Function. *Frontiers in Psychology*, *7*, 745. https://doi.org/10.3389/fpsyg.2016.00745

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working Memory Training Does Not Improve Performance on Measures of Intelligence or Other Measures of "Far Transfer": Evidence From a Meta-Analytic Review [Publisher: SAGE Publications Inc]. *Perspectives on Psychological Science*, *11*(4), 512–534. https://doi.org/10.1177/1745691616635612

Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception [Number: 7397 Publisher: Nature Publishing Group]. *Nature*, *485*(7397), 233–236. https://doi.org/10.1038/nature11020

Michel, C. M., & Brunet, D. (2019). EEG Source Imaging: A Practical Review of the Analysis Steps [Publisher: Frontiers]. *Frontiers in Neurology*, *10*. https://doi.org/10.3389/fneur.2019.00325

Michel, C. M., & He, B. (2019, January). Chapter 6 - EEG source localization. In K. H. Levin & P. Chauvel (Eds.), *Handbook of Clinical Neurology* (pp. 85–101, Vol. 160). Elsevier. https://doi.org/10.1016/B978-0-444-64032-1.00006-0

Michel, C. M., & Koenig, T. (2018). EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: A review. *NeuroImage*, *180*, 577–593. https://doi.org/10.1016/j.neuroimage.2017.11.062

Michel, C. M., Murray, M. M., Lantz, G., Gonzalez, S., Spinelli, L., & Grave de Peralta, R. (2004). EEG source imaging. *Clinical Neurophysiology*, *115*(10), 2195–2222. https://doi.org/10.1016/j.clinph.2004.06.001

Middlebrooks, J. C. (2015, January). Chapter 6 - Sound localization. In M. J. Aminoff, F. Boller, & D. F. Swaab (Eds.), *Handbook of Clinical Neurology* (pp. 99–116, Vol. 129). Elsevier. https://doi.org/10.1016/B978-0-444-62630-1.00006-8

Middlebrooks, J. C. (2017). Spatial Stream Segregation [ISSN: 2197-1897]. In *The Auditory System at the Cocktail Party* (pp. 137–168). Springer, Cham. https://doi.org/10.1007/978-3-319-51662-2_6

Middlebrooks, J. C., Simon, J. Z., Popper, A. N., & Fay, R. R. (Eds.). (2017). *The Auditory System at the Cocktail Party* (Vol. 60). Springer International Publishing. https://doi.org/10.1007/978-3-319-51662-2

Miles, K., McMahon, C., Boisvert, I., Ibrahim, R., de Lissa, P., Graham, P., & Lyxell, B. (2017). Objective Assessment of Listening Effort: Coregistration of Pupillometry and EEG [Publisher: SAGE Publications Inc]. *Trends in Hearing*, *21*, 2331216517706396. https://doi.org/10.1177/2331216517706396

Miller, E. K., & Cohen, J. D. (2001). An Integrative Theory of Prefrontal Cortex Function. *Annual Review of Neuroscience*, *24*(1), 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167

Miyake, A., Emerson, M. J., Padilla, F., & Ahn, J.-c. (2004). Inner speech as a retrieval aid for task goals: The effects of cue type and articulatory suppression in the random task cuing paradigm. *Acta Psychologica*, *115*(2), 123–142. https://doi.org/10.1016/j.actpsy.2003.12.004

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex “Frontal Lobe” Tasks: A Latent Variable Analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Mohammadi, Y., Graversen, C., Manresa, J. B., Østergaard, J., & Andersen, O. K. (2024). Effects of Background Noise and Linguistic Violations on Frontal Theta Oscillations During Effortful Listening. *Ear and Hearing*, *45*(3), 721. https://doi.org/10.1097/AUD.0000000000001464

Mohammadi, Y., Østergaard, J., Graversen, C., Andersen, O. K., & Biurrun Manresa, J. (2023). Validity and reliability of self-reported and neural measures of listening effort. *European Journal of Neuroscience*, *58*(11), 4357–4370. https://doi.org/10.1111/ejn.16187

Moore, B. C. (2008). Basic auditory processes involved in the analysis of speech sounds [Publisher: Royal Society]. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 947–963. https://doi.org/10.1098/rstb.2007.2152

Moore, D. R. (2000). Auditory neuroscience: Is speech special? *Current Biology*, *10*(10), R362–R364. https://doi.org/10.1016/S0960-9822(00)00479-6

Moore, M. J., Byrne, J., Gibson, E. C., Ford, L., & Robinson, G. A. (2024). Hayling and stroop tests tap dissociable deficits and network-level neural correlates [Company: Springer Distributor: Springer Institution: Springer Label: Springer Publisher: Springer Berlin Heidelberg]. *Brain Structure and Function*, *229*(4), 879–896. https://doi.org/10.1007/s00429-024-02767-7

Moore, T. J. (Ed.). (1981). Voice communication jamming research. *Aural Communication in aviation.*

Moore, T. M., & Picou, E. M. (2018). A Potential Bias in Subjective Ratings of Mental Effort. *Journal of Speech, Language, and Hearing Research : JSLHR*, *61*(9), 2405–2421. https://doi.org/10.1044/2018_JSLHR-H-17-0451

Moreau, Q., Parrotta, E., Era, V., Martelli, M. L., & Candidi, M. (2020). Role of the occipito-temporal theta rhythm in hand visual identification [Publisher: American Physiological Society]. *Journal of Neurophysiology*, *123*(1), 167–177. https://doi.org/10.1152/jn.00267.2019

Muncke, J., Kuruvila, I., & Hoppe, U. (2022). Prediction of Speech Intelligibility by Means of EEG Responses to Sentences in Noise [Publisher: Frontiers]. *Frontiers in Neuroscience*, *16*. https://doi.org/10.3389/fnins.2022.876421

Mushtaq, F., Welke, D., Gallagher, A., Pavlov, Y. G., Kouara, L., Bosch-Bayard, J., van den Bosch, J. J. F., Arvaneh, M., Bland, A. R., Chaumon, M., Borck, C., He, X., Luck, S. J., Machizawa, M. G., Pernet, C., Puce, A., Segalowitz, S. J., Rogers, C., Awais, M., . . . Valdes-Sosa, P. (2024). One hundred years of EEG for brain and behaviour research [Publisher: Nature Publishing Group]. *Nature Human Behaviour*, *8*(8), 1437–1443. https://doi.org/10.1038/s41562-024-01941-5

Nagels, L., Gaudrain, E., Vickers, D., Hendriks, P., & Başkent, D. (2021). School-age children benefit from voice gender cue differences for the perception of speech in competing speech. *The Journal of the Acoustical Society of America*, *149*(5), 3328–3344. https://doi.org/10.1121/10.0004791

Neubauer, A. C., & Fink, A. (2009). Intelligence and neural efficiency. *Neuroscience & Biobehavioral Reviews*, *33*(7), 1004–1023. https://doi.org/10.1016/j.neubiorev.2009.04.001

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America*, *95*(2), 1085–1099. https://doi.org/10.1121/1.408469

Norman, D. A., & Shallice, T. (1986). Attention to Action. In *Consciousness and Self-Regulation* (pp. 1–18). Springer, Boston, MA. https://doi.org/10.1007/978-1-4757-0629-1_1

Nourski, K. V., Reale, R. A., Oya, H., Kawasaki, H., Kovach, C. K., Chen, H., Howard, M. A., & Brugge, J. F. (2009). Temporal Envelope of Time-Compressed Speech

Represented in the Human Auditory Cortex. *The Journal of Neuroscience*, *29*(49), 15564–15574. https://doi.org/10.1523/JNEUROSCI.3065-09.2009
Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2010). Broca's Area and Language Processing: Evidence for the Cognitive Control Connection. *Language and Linguistics Compass*, *4*(10), 906–924. https://doi.org/10.1111/j.1749-818X.2010.00244.x
Obleser, J., & Kotz, S. A. (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *NeuroImage*, *55*(2), 713–723. https://doi.org/10.1016/j.neuroimage.2010.12.020
Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., & Maess, B. (2012). Adverse Listening Conditions and Memory Load Drive a Common Alpha Oscillatory Network. *The Journal of Neuroscience*, *32*(36), 12376–12383. https://doi.org/10.1523/JNEUROSCI.4908-11.2012
Onton, J., Delorme, A., & Makeig, S. (2005). Frontal midline EEG dynamics during working memory. *NeuroImage*, *27*(2), 341–356. https://doi.org/10.1016/j.neuroimage.2005.04.014
Organization, W. H. (2021). World Report on Hearing.
O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., Slaney, M., Shamma, S. A., & Lalor, E. C. (2015). Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cerebral Cortex*, *25*(7), 1697–1706. https://doi.org/10.1093/cercor/bht355
Papesh, M. A., Folmer, R. L., & Gallun, F. J. (2017). Cortical Measures of Binaural Processing Predict Spatial Release from Masking Performance. *Frontiers in Human Neuroscience*, *11*. Retrieved November 17, 2022, from https://www.frontiersin.org/articles/10.3389/fnhum.2017.00124
Pascual-Marqui, R. D., Michel, C. M., & Lehmann, D. (1994). Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain. *International Journal of Psychophysiology*, *18*(1), 49–65. https://doi.org/10.1016/0167-8760(84)90014-X
Paul, B. T., Chen, J., Le, T., Lin, V., & Dimitrijevic, A. (2021). Cortical alpha oscillations in cochlear implant users reflect subjective listening effort during speech-in-noise perception (A. Buechner, Ed.). *PLOS ONE*, *16*(7), e0254162. https://doi.org/10.1371/journal.pone.0254162
Pedroni, A., Bahreini, A., & Langer, N. (2019). Automagic: Standardized preprocessing of big EEG data. *NeuroImage*, *200*, 460–473. https://doi.org/10.1016/j.neuroimage.2019.06.046
Peelle, J. E. (2018). Listening Effort: How the Cognitive Consequences of Acoustic Challenge Are Reflected in Brain and Behavior. *Ear and Hearing*, *39*(2), 204–214. https://doi.org/10.1097/AUD.0000000000000494
Peelle, J. E. (2022). How Our Brains Make Sense of Noisy Speech. *Acoustics Today*, *18*(3), 40. https://doi.org/10.1121/AT.2022.18.3.40
Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017
Pernet, C., Rousselet, G., Gaspar, C., & Chauveau, N. (2016). LIMO EEG v1.5 [Accepted: 2016-11-17T12:22:14Z Publisher: University of Edinburgh, Centre for Clinical Brain Sciences]. https://doi.org/10.7488/ds/1557

Pernet, C. R., Appelhoff, S., Gorgolewski, K. J., Flandin, G., Phillips, C., Delorme, A., & Oostenveld, R. (2019). EEG-BIDS, an extension to the brain imaging data structure for electroencephalography [Number: 1 Publisher: Nature Publishing Group]. *Scientific Data*, *6*(1), 103. https://doi.org/10.1038/s41597-019-0104-8

Perrone-Bertolotti, M., Tassin, M., & Meunier, F. (2017). Speech-in-speech perception and executive function involvement (J. Ahveninen, Ed.). *PLOS ONE*, *12*(7), e0180084. https://doi.org/10.1371/journal.pone.0180084

Pfurtscheller, G., Neuper, C., Andrew, C., & Edlinger, G. (1997). Foot and hand area mu rhythms. *International Journal of Psychophysiology*, *26*(1), 121–135. https://doi.org/10.1016/S0167-8760(97)00760-5

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing*, *37*, 5S. https://doi.org/10.1097/AUD.0000000000000312

Power, A. J., Foxe, J. J., Forde, E.-J., Reilly, R. B., & Lalor, E. C. (2012). At what time is the cocktail party? A late locus of selective attention to natural speech. *European Journal of Neuroscience*, *35*(9), 1497–1503. https://doi.org/10.1111/j.1460-9568.2012.08060.x

Pribram, K. H. (1973, January). Chapter 14 - THE PRIMATE FRONTAL CORTEX – EXECUTIVE OF THE BRAIN. In K. H. Pribram & A. R. Luria (Eds.), *Psychophysiology of the Frontal Lobes* (pp. 293–314). Academic Press. https://doi.org/10.1016/B978-0-12-564340-5.50019-6

Puffay, C., Vanthornhout, J., Gillis, M., Clercq, P. D., Accou, B., Hamme, H. V., & Francart, T. (2024). Classifying coherent versus nonsense speech perception from EEG using linguistic speech features [Publisher: Nature Publishing Group]. *Scientific Reports*, *14*(1), 18922. https://doi.org/10.1038/s41598-024-69568-0

Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A.-S., McNamara, J. O., & Williams, S. M. (2001). The Auditory Cortex. In *Neuroscience. 2nd edition*. Sinauer Associates. Retrieved July 30, 2025, from https://www.ncbi.nlm.nih.gov/books/NBK10900/

Qi, Z., Han, M., Wang, Y., de los Angeles, C., Liu, Q., Garel, K., Chen, E. S., Whitfield-Gabrieli, S., Gabrieli, J. D. E., & Perrachione, T. K. (2019). Speech processing and plasticity in the right hemisphere predict variation in adult foreign language learning. *NeuroImage*, *192*, 76–87. https://doi.org/10.1016/j.neuroimage.2019.03.008

Rachana, D., & Neelamegarajan, D. (2024). Development and content validation of coordinate response measure (CRM) corpus in Kannada for informational masking measurement. *The Egyptian Journal of Otolaryngology*, *40*(1), 110. https://doi.org/10.1186/s43163-024-00680-8

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022, December). Robust Speech Recognition via Large-Scale Weak Supervision [arXiv:2212.04356 [eess]]. https://doi.org/10.48550/arXiv.2212.04356

Rennies, J., Best, V., Roverud, E., & Kidd, G. (2019). Energetic and Informational Components of Speech-on-Speech Masking in Binaural Speech Intelligibility and Perceived Listening Effort. *Trends in Hearing*, *23*, 233121651985459. https://doi.org/10.1177/2331216519854597

Rennies, J., Schepker, H., Holube, I., & Kollmeier, B. (2014). Listening effort and speech intelligibility in listening situations affected by noise and reverberation. *The Journal of the Acoustical Society of America*, *136*(5), 2642–2653. https://doi.org/10.1121/1.4897398

Rhebergen, K. S., Versfeld, N. J., & Dreschler, W. A. (2005). Release from informational masking by time reversal of native and non-native interfering speech. *The Journal of the Acoustical Society of America*, *118*(3), 1274–1277. https://doi.org/10.1121/1.2000751

Richter, M. (2016). The Moderating Effect of Success Importance on the Relationship Between Listening Demand and Listening Effort. *Ear and Hearing*, *37*, 111S. https://doi.org/10.1097/AUD.0000000000000295

Rodd, J. M., Johnsrude, I. S., & Davis, M. H. (2010). The role of domain-general frontal systems in language comprehension: Evidence from dual-task interference and semantic ambiguity. *Brain and Language*, *115*(3), 182–188. https://doi.org/10.1016/j.bandl.2010.07.005

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., & Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing [Publisher: Cambridge University Press]. *Applied Psycholinguistics*, *27*(3), 465–485. https://doi.org/10.1017/S014271640606036X

Rönnberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: A framework and a model. *International Journal of Audiology*, *42*(sup1), 68–76. https://doi.org/10.3109/14992020309074626

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, Ö., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*. Retrieved November 29, 2023, from https://www.frontiersin.org/articles/10.3389/fnsys.2013.00031

Ross, J. M., Comstock, D. C., Iversen, J. R., Makeig, S., & Balasubramaniam, R. (2022). Cortical mu rhythms during action and passive music listening [Publisher: American Physiological Society]. *Journal of Neurophysiology*, *127*(1), 213–224. https://doi.org/10.1152/jn.00346.2021

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A., Schulz, J., Fox, P., & Eickhoff, S. (2012). Modelling neural correlates of working memory: A coordinate-based meta-analysis. *Neuroimage*, *60*(1), 830–846. https://doi.org/10.1016/j.neuroimage.2011.11.050

Roushan, H., Gavgani, S. B., & Geravanchizadeh, M. (2023, September). Auditory attention detection in cocktail-party: A microstate study. https://doi.org/10.1101/2023.09.27.559867

Rudner, M., Ng, E. H.-N., Ronnberg, N., Mishra, S., Ronnberg, J., Lunner, T., & Stenfelt, S. (2011). COGNITIVE SPARE CAPACITY AS A MEASURE OF LISTENING EFFORT. *1*(2).

Rudner, M., & Signoret, C. (2016). Editorial: The Role of Working Memory and Executive Function in Communication under Adverse Conditions [Publisher: Frontiers]. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00148

Sala, G., Aksayli, N. D., Tatlidil, K. S., Tatsumi, T., Gondo, Y., & Gobet, F. (2019). Near and Far Transfer in Cognitive Training: A Second-Order Meta-Analysis (R. Zwaan & P. Verkoeijen, Eds.). *Collabra: Psychology*, *5*(1), 18. https://doi.org/10.1525/collabra.203

Sala, G., & Gobet, F. (2017). Does Far Transfer Exist? Negative Evidence From Chess, Music, and Working Memory Training [Publisher: SAGE Publications Inc]. *Current Directions in Psychological Science*, *26*(6), 515–520. https://doi.org/10.1177/0963721417712760

Scannella, S., Peysakhovich, V., Ehrig, F., Lepron, E., & Dehais, F. (2018). Assessment of Ocular and Physiological Metrics to Discriminate Flight Phases in Real Light Aircraft [Publisher: SAGE Publications Inc]. *Human Factors*, *60*(7), 922–935. https://doi.org/10.1177/0018720818787135

Scott, S. K., & McGettigan, C. (2013). The neural processing of masked speech. *Hearing Research*, *303*, 58–66. https://doi.org/10.1016/j.heares.2013.05.001

Seifi Ala, T., Graversen, C., Wendt, D., Alickovic, E., Whitmer, W. M., & Lunner, T. (2020). An exploratory Study of EEG Alpha Oscillation and Pupil Dilation in Hearing-Aid Users During Effortful listening to Continuous Speech (I. Yasin, Ed.). *PLOS ONE*, *15*(7), e0235782. https://doi.org/10.1371/journal.pone.0235782

Semeraro, H. D., Rowan, D., van Besouw, R. M., & Allsopp, A. A. (2017). Development and evaluation of the British English coordinate response measure speech-in-noise test as an occupational hearing assessment tool [Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/14992027.2017.1317370]. *International Journal of Audiology*, *56*(10), 749–758. https://doi.org/10.1080/14992027.2017.1317370

Shields, C., Willis, H., Nichani, J., Sladen, M., & Kluk-de Kort, K. (2022). Listening effort: WHAT is it, HOW is it measured and WHY is it important? [Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/14670100.2021.1992941]. *Cochlear Implants International*, *23*(2), 114–117. https://doi.org/10.1080/14670100.2021.1992941

Shilling, V. M., Chetwynd, A., & Rabbitt, P. M. A. (2002). Individual inconsistency across measures of inhibition: An investigation of the construct validity of inhibition in older adults. *Neuropsychologia*, *40*(6), 605–619. https://doi.org/10.1016/S0028-3932(01)00157-9

Shinn-Cunningham, B., Best, V., & Lee, A. K. C. (2017). Auditory Object Formation and Selection [ISSN: 2197-1897]. In *The Auditory System at the Cocktail Party* (pp. 7–40). Springer, Cham. https://doi.org/10.1007/978-3-319-51662-2_2

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do "Brain-Training" Programs Work? [Publisher: SAGE Publications Inc]. *Psychological Science in the Public Interest*, *17*(3), 103–186. https://doi.org/10.1177/1529100616661983

Smiljanić, R., & Bradlow, A. R. (2011). Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentedness. *The Journal of the Acoustical Society of America*, *130*(6), 4020–4031. https://doi.org/10.1121/1.3652882

Stenbäck, V., Hällgren, M., & Larsby, B. (2016). Executive functions and working memory capacity in speech communication under adverse conditions.

Speech, Language and Hearing, *19*(4), 218–226. https://doi.org/10.1080/2050571X.2016.1196034

Stenbäck, V., Marsja, E., H, ä. M., Lyxell, B., & Larsby, B. (2021). The Contribution of Age, Working Memory Capacity, and Inhibitory Control on Speech Recognition in Noise in Young and Older Adult Listeners [Publisher: American Speech-Language-Hearing Association]. *Journal of Speech, Language, and Hearing Research*, *64*(11), 4513–4523. https://doi.org/10.1044/2021_JSLHR-20-00251

Strand, J. F., Brown, V. A., Merchant, M. B., Brown, H. E., & Smith, J. (2018). Measuring Listening Effort: Convergent Validity, Sensitivity, and Links With Cognitive and Personality Measures. *Journal of Speech, Language, and Hearing Research*, *61*(6), 1463–1486. https://doi.org/10.1044/2018_JSLHR-H-17-0257

Strauß, A., Wöstmann, M., & Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition [Publisher: Frontiers]. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00350

Strauss, D. J., & Francis, A. L. (2017). Toward a taxonomic model of attention in effortful listening. *Cognitive, Affective, & Behavioral Neuroscience*, *17*(4), 809–825. https://doi.org/10.3758/s13415-017-0513-0

Strobach, T., & Karbach, J. (Eds.). (2021). *Cognitive Training: An Overview of Features and Applications*. Springer International Publishing. https://doi.org/10.1007/978-3-030-39292-5

Stroop, J. R. (1935). Studies of interference in serial verbal reactions [Place: US Publisher: Psychological Review Company]. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Studebaker, G. A. (1985). A "Rationalized" Arcsine Transform [Publisher: American Speech-Language-Hearing Association]. *Journal of Speech, Language, and Hearing Research*, *28*(3), 455–462. https://doi.org/10.1044/jshr.2803.455

Stuss, D. T. (2011). Functions of the Frontal Lobes: Relation to Executive Functions. *Journal of the International Neuropsychological Society*, *17*(5), 759–765. https://doi.org/10.1017/S1355617711000695

Stuss, D. T., & Alexander, M. P. (2000). Executive functions and the frontal lobes: A conceptual view [Company: Springer Distributor: Springer Institution: Springer Label: Springer Number: 3 Publisher: Springer-Verlag]. *Psychological Research*, *63*(3), 289–298. https://doi.org/10.1007/s004269900007

Stuss, D. T., & Beson, F. D. (1984). Neuropsychological studies of the frontal lobes. *Psychological Bulletin*. Retrieved May 23, 2025, from https://www.academia.edu/1498172/neuropsych_studies_of_the_frontal_lobes

Sur, S., & Sinha, V. (2009). Event-related potential: An overview [Num Pages: 3 Place: Mumbai, Mumbai Publisher: Medknow Publications & Media Pvt. Ltd.], *18*(1), 70–73. https://doi.org/10.4103/0972-6748.57865

Taatgen, N. A. (2021). Theoretical Models of Training and Transfer Effects. In *Cognitive Training* (pp. 41–54). Springer, Cham. https://doi.org/10.1007/978-3-030-39292-5_4

Tarailis, P., Koenig, T., Michel, C. M., & Griškova-Bulanova, I. (2024). The Functional Aspects of Resting EEG Microstates: A Systematic Review. *Brain Topography*, *37*(2), 181–217. https://doi.org/10.1007/s10548-023-00958-9

Teng, X., & Poeppel, D. (2020). Theta and Gamma Bands Encode Acoustic Dynamics over Wide-Ranging Timescales. *Cerebral Cortex*, *30*(4), 2600–2614. https://doi.org/10.1093/cercor/bhz263

Thompson, E. R., Iyer, N., Simpson, B. D., Wakefield, G. H., Kieras, D. E., & Brungart, D. S. (2015). Enhancing listener strategies using a payoff matrix in speech-on-speech masking experimentsa). *The Journal of the Acoustical Society of America*, *138*(3), 1297–1304. https://doi.org/10.1121/1.4928395

Thorell, L. B., Lindqvist, S., Bergman Nutley, S., Bohlin, G., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science*, *12*(1), 106–113. https://doi.org/10.1111/j.1467-7687.2008.00745.x

Tian, Y., Xu, W., & Yang, L. (2018). Cortical Classification with Rhythm Entropy for Error Processing in Cocktail Party Environment Based on Scalp EEG Recording [Number: 1 Publisher: Nature Publishing Group]. *Scientific Reports*, *8*(1), 6070. https://doi.org/10.1038/s41598-018-24535-4

Tremblay, K. L., Pinto, A., Fischer, M. E., Klein, B. E. K., Klein, R., Levy, S., Tweed, T. S., & Cruickshanks, K. J. (2015). Self-Reported Hearing Difficulties Among Adults With Normal Audiograms: The Beaver Dam Offspring Study. *Ear and Hearing*, *36*(6), e290–299. https://doi.org/10.1097/AUD.0000000000000195

van Veen, V., Cohen, J. D., Botvinick, M. M., Stenger, V. A., & Carter, C. S. (2001). Anterior Cingulate Cortex, Conflict Monitoring, and Levels of Processing. *NeuroImage*, *14*(6), 1302–1308. https://doi.org/10.1006/nimg.2001.0923

Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *The Journal of the Acoustical Society of America*, *121*(1), 519–526. https://doi.org/10.1121/1.2400666

Van Hedger, S. C., & Johnsrude, I. S. (2022). Speech Perception Under Adverse Listening Conditions [ISSN: 2197-1897]. In *Speech Perception* (pp. 141–171). Springer, Cham. https://doi.org/10.1007/978-3-030-81542-4_6

Verbruggen, F., Adams, R., & Chambers, C. D. (2012). Proactive Motor Control Reduces Monetary Risk Taking in Gambling. *Psychological Science*, *23*(7), 805–815. https://doi.org/10.1177/0956797611434538

Verbruggen, F., Liefooghe, B., & Vandierendonck, A. (2004). The interaction between stop signal inhibition and distractor interference in the flanker and Stroop task. *Acta Psychologica*, *116*(1), 21–37. https://doi.org/10.1016/j.actpsy.2003.12.011

Wagener, K., Josvassen, J. L., & Ardenkjær, R. (2003). Design, optimization and evaluation of a Danish sentence test in noise: Diseño, optimización y evaluación de la prueba Danesa de frases en ruido. *International Journal of Audiology*, *42*(1), 10–17. https://doi.org/10.3109/14992020309056080

Wagener, K., Brand, T., & Kollmeier, B. (1999). Entwicklung und evaluation eines satztests für die deutsche sprache I-III: Design, optimierung und evaluation des oldenburger satztests. *Zeitschrift für Audiologie*.

Wager, T. D., Jonides, J., & Reading, S. (2004). Neuroimaging studies of shifting attention: A meta-analysis. *NeuroImage*, *22*(4), 1679–1693. https://doi.org/10.1016/j.neuroimage.2004.03.052

Walter, W. G. (1936). The location of cerebral tumors by electroencephalography. *The Lancet*, *228*(5893), 305–308. https://doi.org/10.1016/S0140-6736(01)05173-X

Wang, Y., Lu, Z., Yang, X., & Liu, C. (2019). Measuring Mandarin Speech Recognition Thresholds Using the Method of Adaptive Tracking [Publisher: American Speech-Language-Hearing Association]. *Journal of Speech, Language, and Hearing Research*, *62*(6), 2009–2017. https://doi.org/10.1044/2019_JSLHR-H-18-0162

Warzybok, A., Brand, T., Wagener, K. C., & Kollmeier, B. (2015). How much does language proficiency by non-native listeners influence speech audiometric tests in noise? *International Journal of Audiology*, *54*(sup2), 88–99. https://doi.org/10.3109/14992027.2015.1063715

Warzybok, A., Zokoll, M., Wardenga, N., Ozimek, E., Boboshko, M., & Kollmeier, B. (2015). Development of the Russian matrix sentence test [Publisher: Informa UK Limited]. *International Journal of Audiology*, *54*(sup2), 35–43. https://doi.org/10.3109/14992027.2015.1020969

Weisz, N., Hartmann, T., Müller, N., Lorenz, I., & Obleser, J. (2011). Alpha Rhythms in Audition: Cognitive and Clinical Perspectives. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00073

Weisz, N., & Obleser, J. (2014). Synchronisation signatures in the listening brain: A perspective from non-invasive neuroelectrophysiology. *Hearing Research*, *307*, 16–28. https://doi.org/10.1016/j.heares.2013.07.009

Whitton, J. P., Hancock, K. E., & Polley, D. B. (2014). Immersive audiomotor game play enhances neural and perceptual salience of weak signals in noise. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(25), E2606–E2615. https://doi.org/10.1073/pnas.1322184111

Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful Listening: The Processing of Degraded Speech Depends Critically on Attention [Publisher: Society for Neuroscience Section: Articles]. *Journal of Neuroscience*, *32*(40), 14010–14021. https://doi.org/10.1523/JNEUROSCI.1528-12.2012

Wilsch, A., Henry, M. J., Herrmann, B., Maess, B., & Obleser, J. (2015). Alpha Oscillatory Dynamics Index Temporal Expectation Benefits in Working Memory. *Cerebral Cortex*, *25*(7), 1938–1946. https://doi.org/10.1093/cercor/bhu004

Winn, M. B., & Teece, K. H. (2021). Listening Effort Is Not the Same as Speech Intelligibility Score. *Trends in Hearing*, *25*, 233121652110276. https://doi.org/10.1177/23312165211027688

Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best Practices and Advice for Using Pupillometry to Measure Listening Effort: An Introduction for Those Who Want to Get Started [Publisher: SAGE Publications Inc]. *Trends in Hearing*, *22*, 2331216518800869. https://doi.org/10.1177/2331216518800869

Wisniewski, M. G. (2017). Indices of Effortful Listening Can Be Mined from Existing Electroencephalographic Data. *Ear & Hearing*, *38*(1), e69–e73. https://doi.org/10.1097/AUD.0000000000000354

Wisniewski, M. G., Iyer, N., Thompson, E. R., & Simpson, B. D. (2018). Sustained frontal midline theta enhancements during effortful listening track working memory demands. *Hearing Research*, *358*, 37–41. https://doi.org/10.1016/j.heares.2017.11.009

Wisniewski, M. G., Joyner, C. N., Zakrzewski, A. C., & Anguiano, A. (2023). Learning to detect auditory signals in noise: Active top-down selection and stable change in signal representations. *Journal of Experimental Psychology: Hu-*

man Perception and Performance, *49*(3), 428–440. https://doi.org/10.1037/xhp0001082

Wisniewski, M. G., Joyner, C. N., Zakrzewski, A. C., & Makeig, S. (2024). Finding tau rhythms in EEG: An independent component analysis approach. *Human Brain Mapping*, *45*(2), e26572. https://doi.org/10.1002/hbm.26572

Wisniewski, M. G., Thompson, E. R., & Iyer, N. (2017). Theta- and alpha-power enhancements in the electroencephalogram as an auditory delayed match-to-sample task becomes impossibly difficult. *Psychophysiology*, *54*(12), 1916–1928. https://doi.org/10.1111/psyp.12968

Wisniewski, M. G., Thompson, E. R., Iyer, N., Estepp, J. R., Goder-Reiser, M. N., & Sullivan, S. C. (2015). Frontal midline ϑ power as an index of listening effort. *NeuroReport*, *26*(2), 94–99. https://doi.org/10.1097/WNR.0000000000000306

Wisniewski, M. G., & Zakrzewski, A. C. (2023). Effortful Listening Produces Both Enhancement and Suppression of Alpha in the EEG. *Auditory Perception & Cognition*, 1–11. https://doi.org/10.1080/25742442.2023.2218239

Wisniewski, M. G., Zakrzewski, A. C., Bell, D. R., & Wheeler, M. (2021). EEG power spectral dynamics associated with listening in adverse conditions. *Psychophysiology*, *58*(9). https://doi.org/10.1111/psyp.13877

Wong, L. L. N., & Soli, S. D. (2005). Development of the Cantonese Hearing In Noise Test (CHINT). *Ear and Hearing*, *26*(3), 276. Retrieved July 10, 2025, from https://journals.lww.com/ear-hearing/abstract/2005/06000/development_of_the_cantonese_hearing_in_noise_test.4.aspx

Ye, Z., & Zhou, X. (2009). Conflict control during sentence comprehension: fMRI evidence. *NeuroImage*, *48*(1), 280–290. https://doi.org/10.1016/j.neuroimage.2009.06.032

Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R. (2006). Top–down and bottom–up processes in speech comprehension. *NeuroImage*, *32*(4), 1826–1836. https://doi.org/10.1016/j.neuroimage.2006.04.199

Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The Pupil Dilation Response to Auditory Stimuli: Current State of Knowledge [Publisher: SAGE Publications Inc]. *Trends in Hearing*, *22*, 2331216518777174. https://doi.org/10.1177/2331216518777174

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil Response as an Indication of Effortful Listening: The Influence of Sentence Intelligibility. *Ear and Hearing*, *31*(4), 480. https://doi.org/10.1097/AUD.0b013e3181d4f251

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive Load During Speech Perception in Noise: The Influence of Age, Hearing Loss, and Cognition on the Pupil Response. *Ear and Hearing*, *32*(4), 498. https://doi.org/10.1097/AUD.0b013e31820512bb

Zekveld, A. A., Rudner, M., Kramer, S. E., Lyzenga, J., & Rönnberg, J. (2014). Cognitive processing load during listening is reduced more by decreasing voice similarity than by increasing spatial separation between target and masker speech. *Frontiers in Neuroscience*, *8*, 88. https://doi.org/10.3389/fnins.2014.00088

Zhang, R., Geng, X., & Lee, T. M. C. (2017). Large scale functional neural network correlates of response inhibition: An fMRI meta-analysis. *Brain Structure and Function*, *222*(9), 3973–3990. https://doi.org/10.1007/s00429-017-1443-x

Zhao, X., Chen, L., & Maes, J. H. (2018). Training and transfer effects of response inhibition training in children and adults. *Developmental Science*, *21*(1), e12511. https://doi.org/10.1111/desc.12511

Zheng, Y., & Guan, J. (2018). Cochlear Synaptopathy: A Review of Hidden Hearing Loss.

Zimpfer, V., Andéol, G., Blanck, G., Suied, C., & Fux, T. (2020). Development of a French version of the Modified Rhyme Test. *The Journal of the Acoustical Society of America*, *147*(1), EL55–EL61. https://doi.org/10.1121/10.0000559

Zink, N., Lenartowicz, A., & Markett, S. (2021). A new era for executive function research: On the transition from centralized to distributed executive functioning. *Neuroscience & Biobehavioral Reviews*, *124*, 235–244. https://doi.org/10.1016/j.neubiorev.2021.02.011

Zue, V., Seneff, S., & Glass, J. (1990). Speech database development at MIT: Timit and beyond. *Speech Communication*, *9*(4), 351–356. https://doi.org/10.1016/0167-6393(90)90010-7

VI

Appendix

Appendix

1 Audiometry of each Study

Study	All Frequencies	< 2500 Hz	2500 Hz
First Study (n=51)	-.3 ± 8.2	-1.1 ± 6.8	4.9 ± 3.6
Second Study (=30)	-.9 ± 8.3	-1.9 ± 6.7	6.3 ± 13.7
Third Study (n=60)	1.5 ± 10.6	1.0 ± 9.9	4.7 ± 12.4

Table 10.1: Audiometric measure for each study. Mean ± sd hearing level for all frequencies (.25, .5, 1, 2, 4, 6, 8 kHz and 12.5 kHz), medium frequencies (< 2500 Hz) and high frequencies (2500 Hz)

2 Second Study

2.1 Power Calculation for the second study

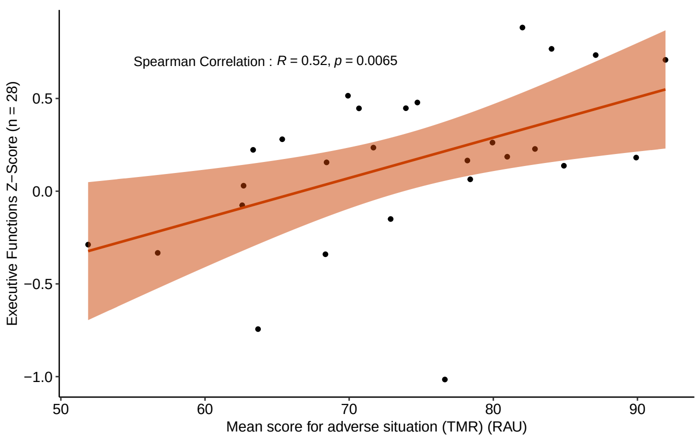


Figure 10.1: Correlation between EF Z-scores and scores and intelligibility in adverse situation (TMR). Calculated on 28 participants from the first study that performed the EF tasks in an independent external study?

2.2 IC clusters ERSPs per TMR/SNR

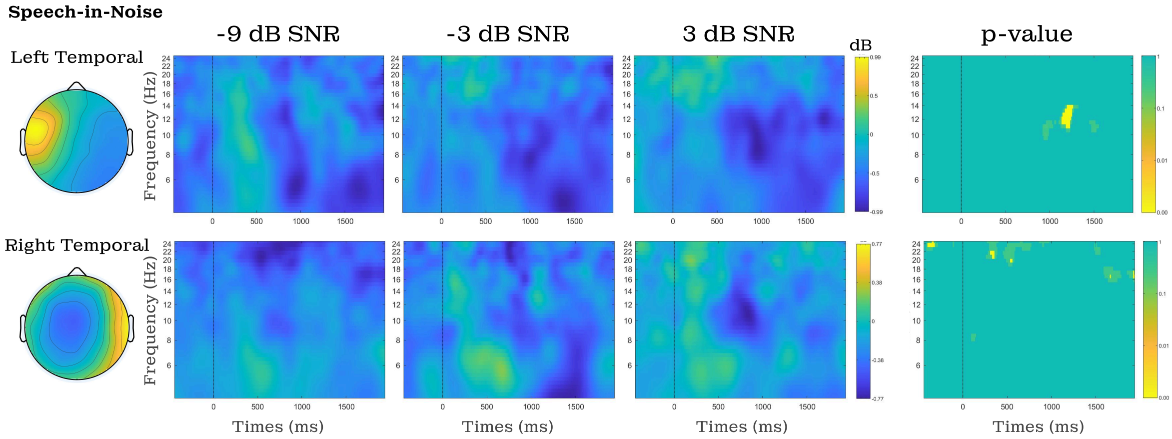


Figure 10.2: ERSPs for each SNR condition for the left and right temporal independent component clusters in the SIN condition.

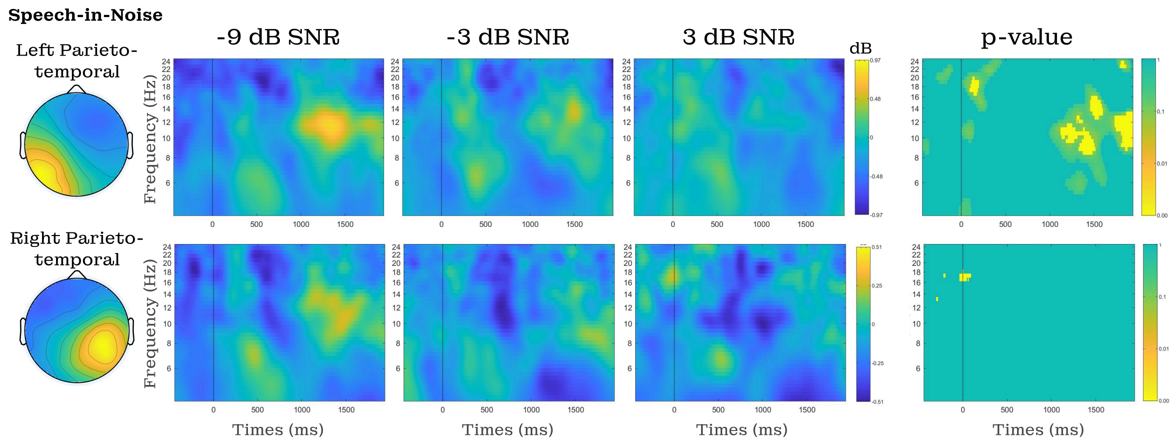


Figure 10.3: ERSPs for each SNR condition for the left and right parieto-temporal independent component clusters in the SIN condition.

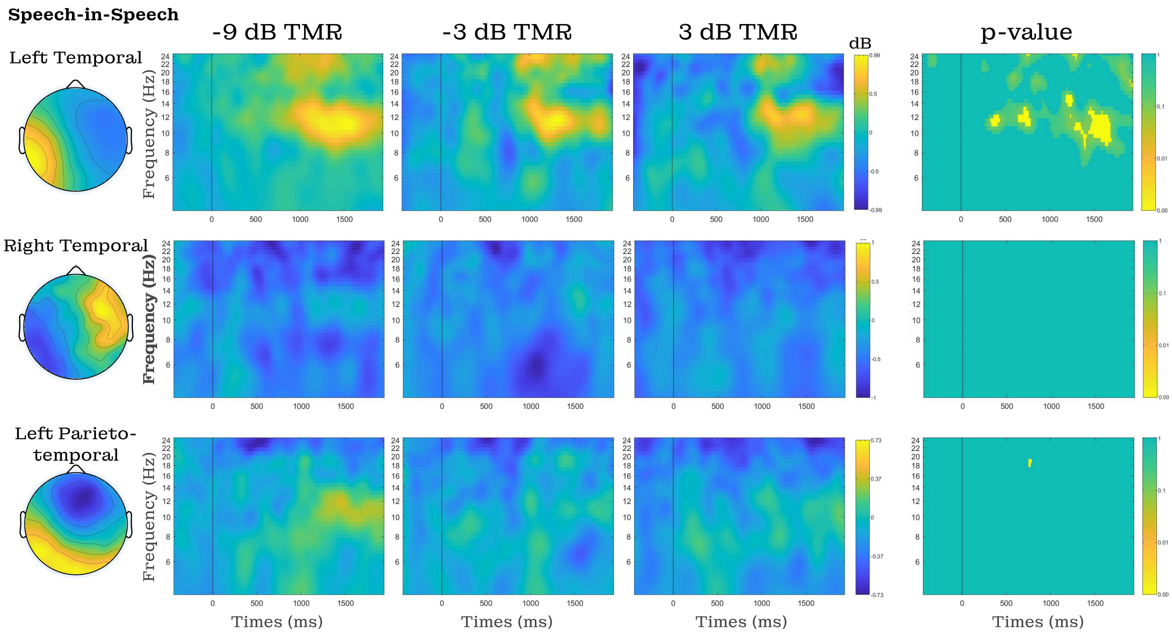


Figure 10.4: ERSPs for each SNR condition for the left parieto-temporal and left and right temporal independent component clusters in the SIS condition.

2.3 Executive Functions and Microstates in SIN and SIS

Microstates

Microstates with 4 to 7 clusters (k) were extracted using the microstates pipeline described in Section 3.2 in the Chapter 5, the prototypical maps are presented in Figure 10.5.

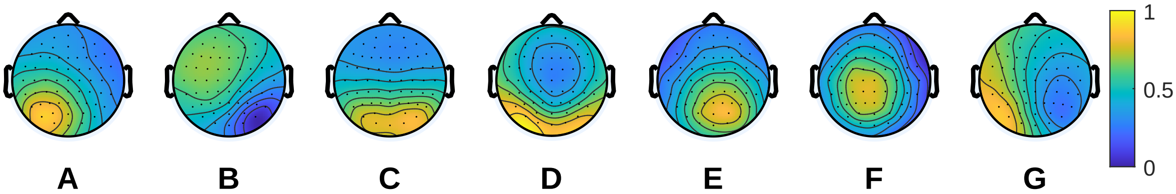


Figure 10.5: Microstates topographical group maps for k =7.

Metrics

The microstates metrics for closed eyes resting state and k=7 are presented on the Table 10.2.

MS	Coverage	Duration	Occurrence
A	1.78 ± .49	60.19 ± 5.79	.11 ± .03
B	2.44 ± .6	67.86 ± 10.53	.17 ± .07
C	3.22 ± .45	78.11 ± 12.99	.25 ± .07
D	2.36 ± .67	64.38 ± 7.1	.16 ± .06
E	2.31 ± .7	65.34 ± 9	.15 ± .06
F	1.22 ± .51	54.63 ± 4.29	.07 ± .03
G	1.53 ± .42	58.21 ± 3.69	.09 ± .02

Table 10.2: Microstates (MS) mean and standard deviation of metrics for $k = 7$

Correlations with Speech Intelligibility

Coverage The Pearson correlations between microstates coverage and SI for each SIN and SIS level level are presented in Figure 10.6, with the p-value and the r reported for each correlation.

Significant correlations were observed between SI in SIN condition at -3 dB SNR and coverage of microstates C ($p < .01, r = -.55$) and G ($p = .04, r = .39$). No other significant correlations were observed.

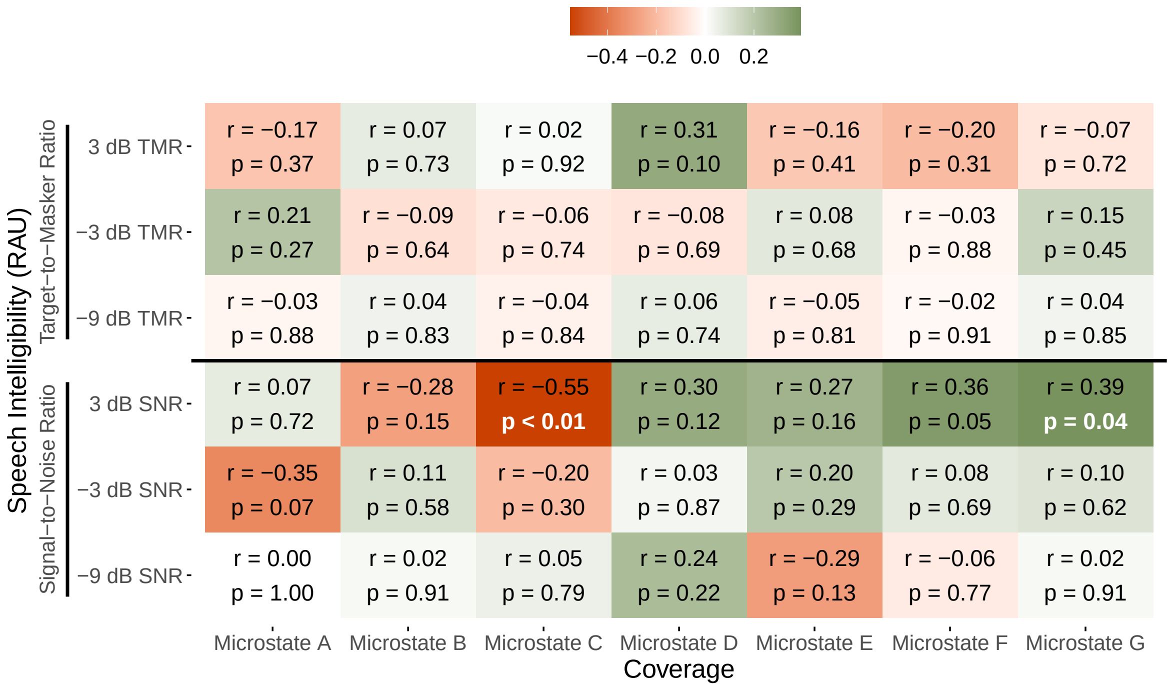


Figure 10.6: Pearson correlations between Microstates (k=7) coverage and SI (RAU) in SIN and SIS. In white, the significant correlations.

Occurrence The Pearson correlations between microstates occurrences and SI for each SIN and SIS level are presented in Figure 10.7, with the p-value and the r reported for each correlation.

Significant correlations were observed between SI in SIN condition at 3 dB SNR and occurrence of microstates C ($p < .01, r = -.55$) and F ($p = .04, r = .38$). No other significant correlations were observed.

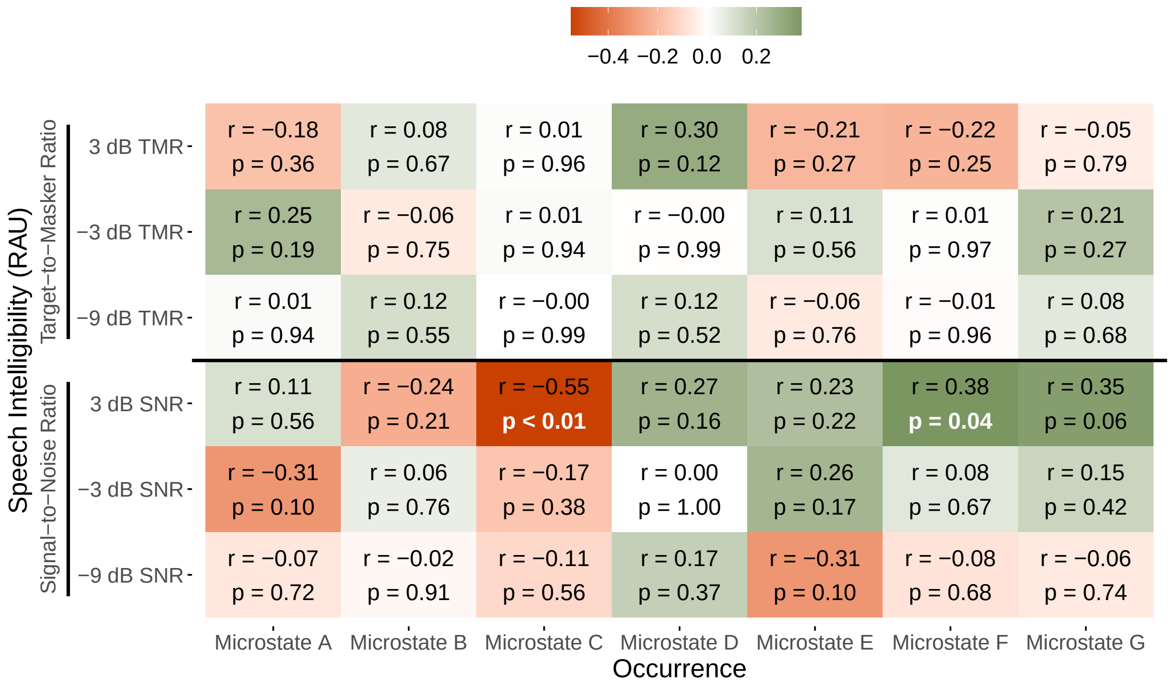


Figure 10.7: Pearson correlations between Microstates (k=7) occurrence and SI (RAU) in SIN and SIS. In white, the significant correlations.

Duration The Pearson correlations between microstates durations and SI for each SIN and SIS level are presented in Figure 10.8, with the p-value and the r reported for each correlation.

A significant correlation was observed between SI in SIN condition at 3 dB SNR and duration of microstates C ($p = .04, r = -.38$) and microstates D ($p = .04, r = .38$). At -9dB SNR, a significant correlation was observed with the microstates D ($p = .01, r = .46$) and the microstates G($p = .04, r = .39$). No other significant correlations were observed.

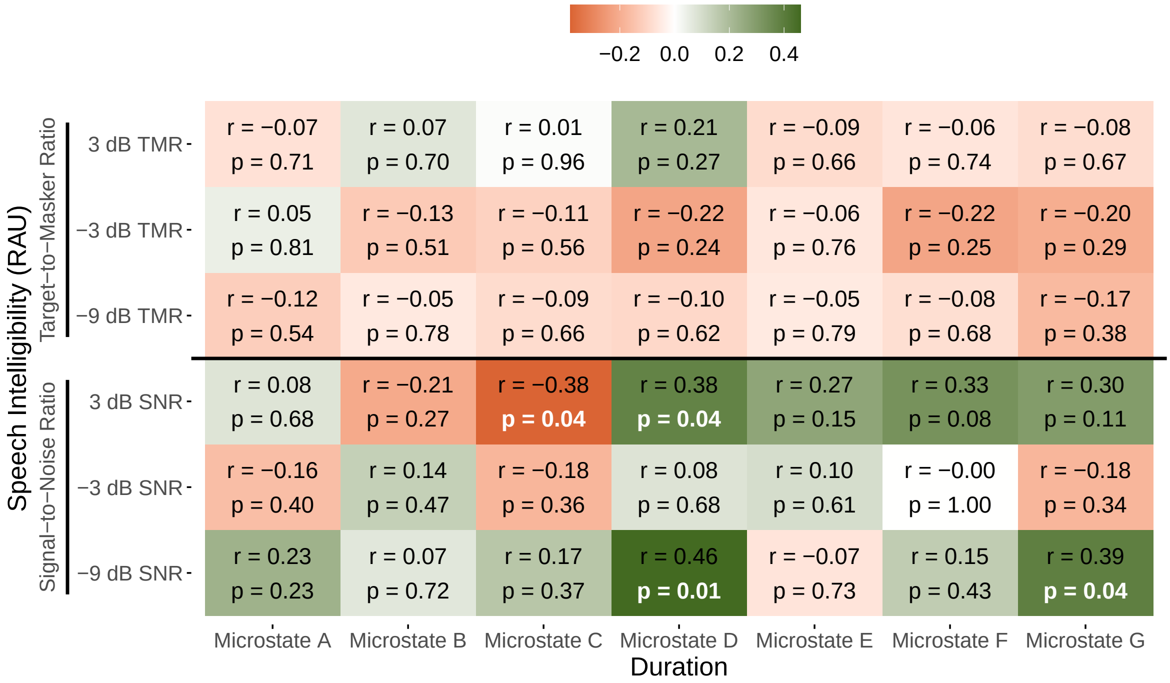


Figure 10.8: Pearson correlations between Microstates (k=7) duration and SI (RAU) in SIN and SIS. In white, the significant correlations.

Correlations with Listening Effort

The Pearson correlations between microstates metrics at k=7 and LE for SIS and SIN showed significant results between the coverage of microstates A and LE at - 9 dB TMR ($p = .04, r = -.38$), and between the duration of microstates A and LE at - 9 dB SNR ($p = .04, r = -.39$) as well as 3 dB TMR ($p = .04, r = -.38$). No other significant results were observed.

3 Third Study

3.1 Inhibition Training

Comparison of Arrow Stroop performances after one repetition of the task between the control and the inhibition-trained group. To do so, we compared the time reactions of the conditions (congruent, incongruent, and control) using the first session performances of both groups and the post-training session for the control group and the first training session for the inhibition-trained group. The three-way ANOVA (see Table 10.3) revealed no interaction between group $\times$ session $\times$ condition. A group $\times$ session interaction was observed and is represented in Figure 10.9 with post-hoc results.

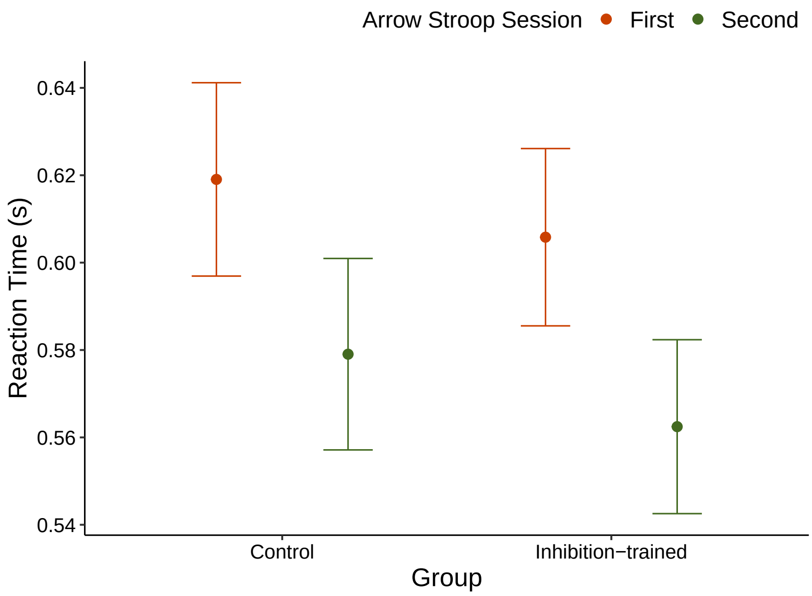


Figure 10.9: Mean Arrow Stroop reaction times (s) with 95% confidence intervals for all Arrow Stroop condition (congruent, incongruent, control), in control and inhibition-trained groups, across the first and the second session of the Arrow Stroop. For the control group, the second session corresponds to the post-training session, for the inhibition-trained group, to the first training session (Day 3).

	Df	F	η_p^2	$p - value$
Group	1	.44	.01	.5127
Ses	1	39.3	.51	$< .001$
Group:Ses	1	.06	.00	.8019
Condition	2	141.6	.79	$< .001$
Group:Condition	2	.22	.01	.76
Ses:Condition	2	2.9	.07	**.0640**
Group:Ses:Condition	2	.08	.00	.92

Table 10.3: Three-way ANOVA on reaction times, with Arrow Stroop Condition (congruent, incongruent, control), Session (first *vs.* second Arrow Stroop session), and Group (inhibition-trained *vs.* control), as factors.

3.2 Training - Speech Intelligibility

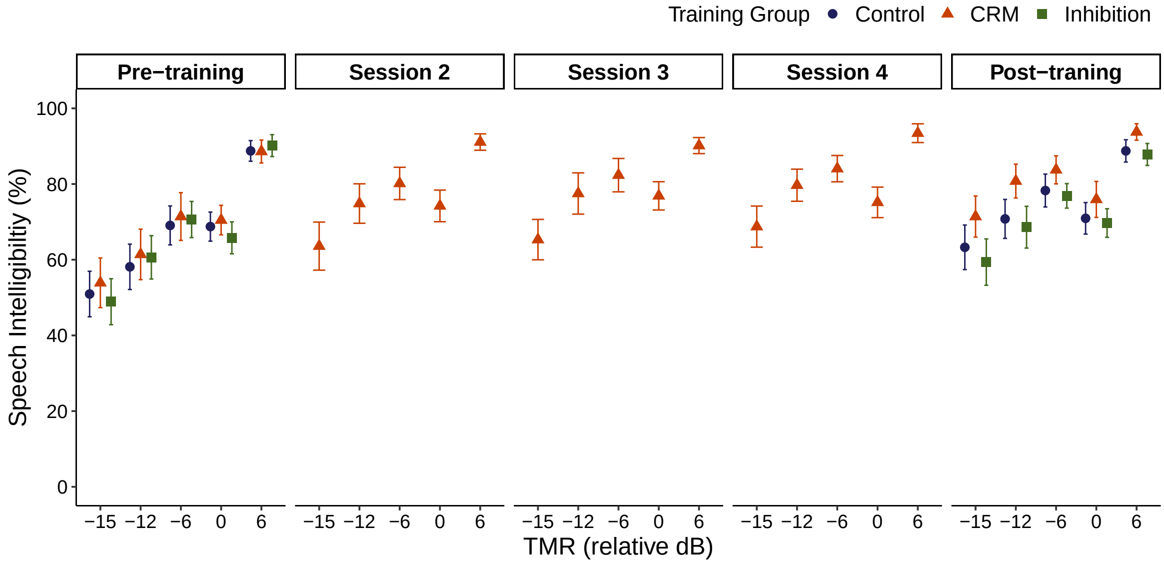


Figure 10.10: Mean speech intelligibility (% correct responses) with 95% confidence intervals for the three training groups (Control: blue; Inhibition: green; CRM: orange) across each session.

3.3 Training - Subjective Listening Effort

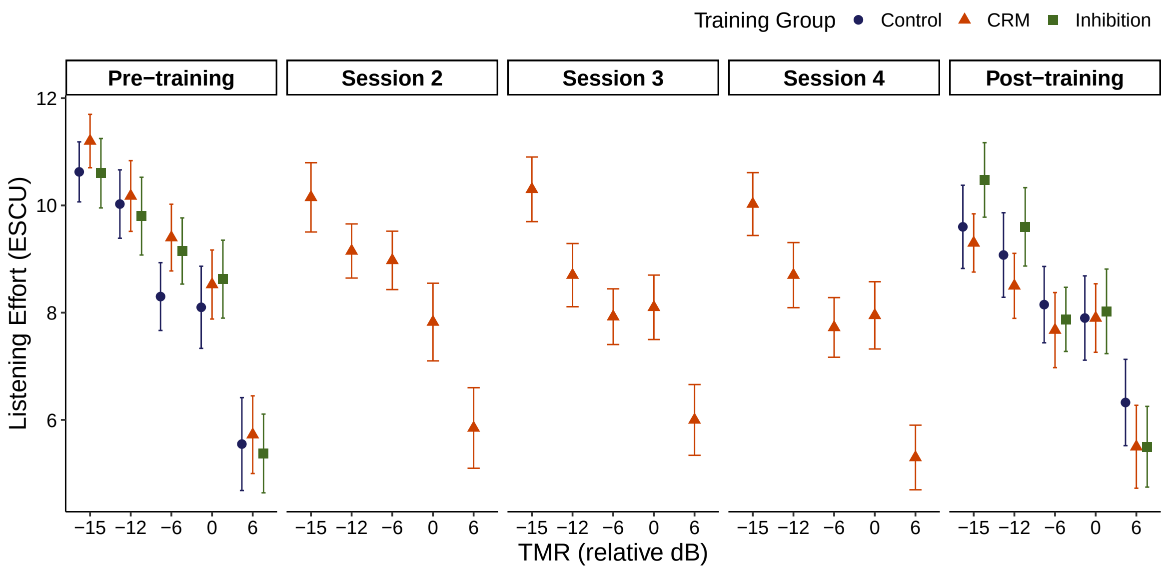


Figure 10.11: Mean subjective listening effort (ESCU) with 95% confidence intervals for the three training groups (Control: blue; Inhibition: green; CRM: orange) across each session.

3.4 Cognitive Training Time Frequency Analyses

Pre-Training Topographies and ERSPs per TMR per Group

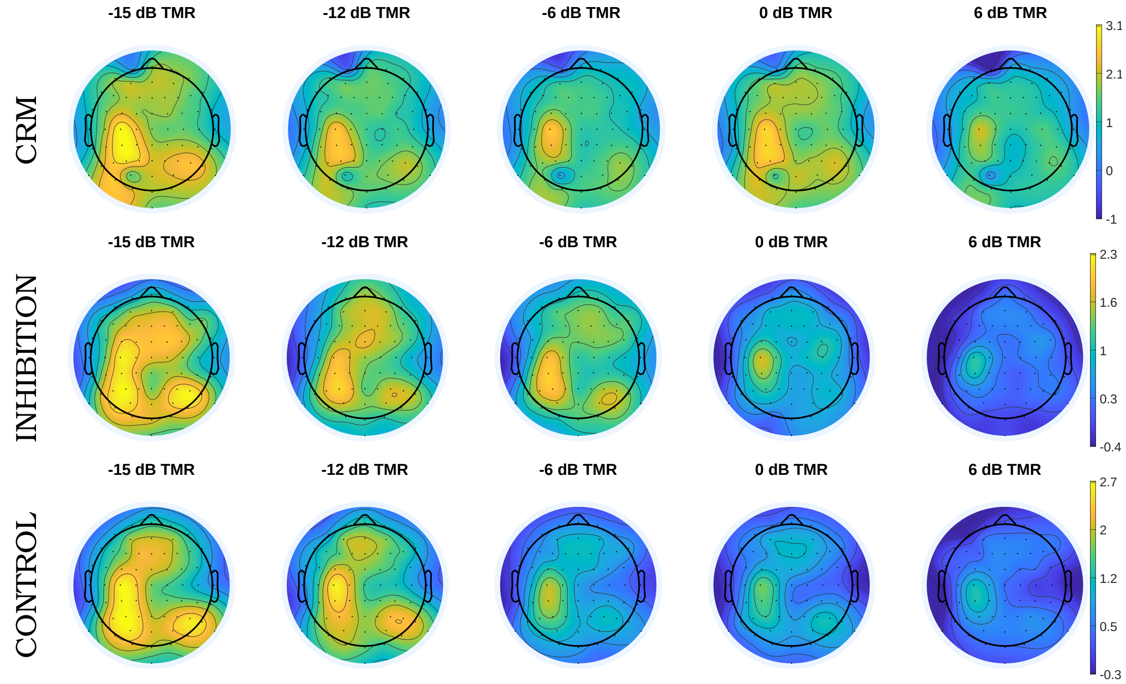


Figure 10.12: Topographical maps of mean alpha power (8–12 Hz; baseline-corrected) over the 0–2000 ms time window (0 = CRM sentence onset) in pre-training session, for each group. The right panels show the p-value map indicating significant differences between TMRs.

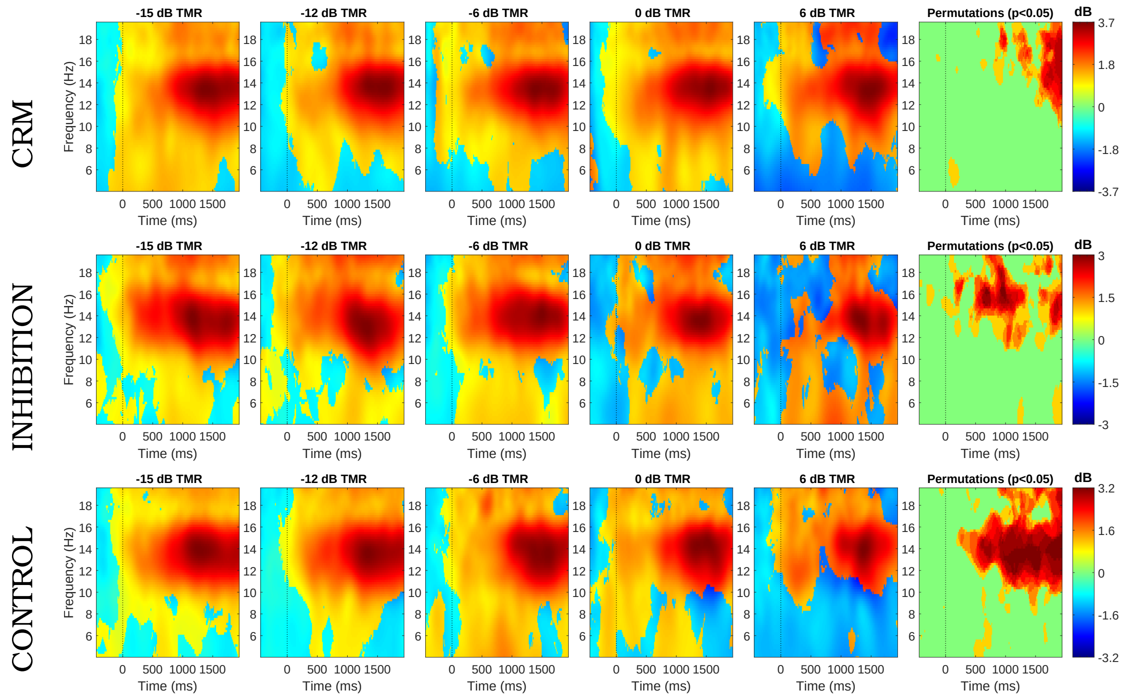


Figure 10.13: Time–frequency representations (ERSPs) showing the Session × Group (Control, CRM) interaction in the ROI (T7, C5, C3, P7, P5, P3, TP7, CP5, and CP3) in the pre-training session, for each group. The right panels show the p-value map indicating significant differences between TMRs

3.5 Microstates

Maps

Microstates prototypical maps, on the session level, for closed eyes resting state with 7 clusters (k) were extracted using the microstates pipeline described in (METHOD) on the session and group levels. The prototypical maps of the study are presented in Figure 10.14.

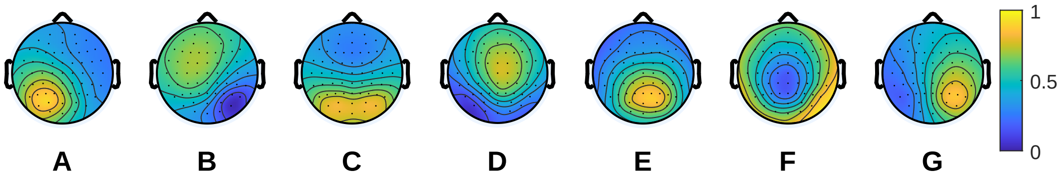


Figure 10.14: Prototypical maps (k=7) MS for the whole Study

Metrics

The microstates metrics for closed eyes resting state and k=7 are presented on the Tables 10.4 and 10.5.

MS	Occurrence	Duration	Coverage
A	1.79 ± .51	63.5 ± 7.4	.11 ± .04
B	2.36 ± .59	69.98 ± 13.17	.17 ± .06
C	3.19 ± .48	87.61 ± 19.81	.28 ± .08
D	1.86 ± .67	63.18 ± 8.98	.12 ± .06
E	2.16 ± .59	68.53 ± 11.47	.15 ± .06
F	1.25 ± .42	57.1 ± 5.63	.07 ± .03
G	1.52 ± .52	60.7 ± 7.48	.09 ± .04

Table 10.4: Microstates metrics of Pre-training Session (mean ± SD) for k = 7

MS	Occurrence	Duration	Coverage
A	1.82 ± .49	63.58 ± 7.71	.12 ± .04
B	2.25 ± .53	68.64 ± 10.92	.16 ± .05
C	3.17 ± .49	86.1 ± 18.62	.28 ± .08
D	1.97 ± .62	64.25 ± 8.53	.13 ± .05
E	2.18 ± .62	68.94 ± 11.75	.15 ± .06
F	1.37 ± .48	58.78 ± 7.39	.08 ± .03
G	1.46 ± .53	59.25 ± 6.96	.09 ± .04

Table 10.5: Microstates metrics of Post-training Session (mean ± SD) for k = 7

No first-order $group \times session$ interactions were observed when computing three-way ANOVA for each metric (Coverage, Duration and Occurrence). The ANOVA tables can be found in Annexes.

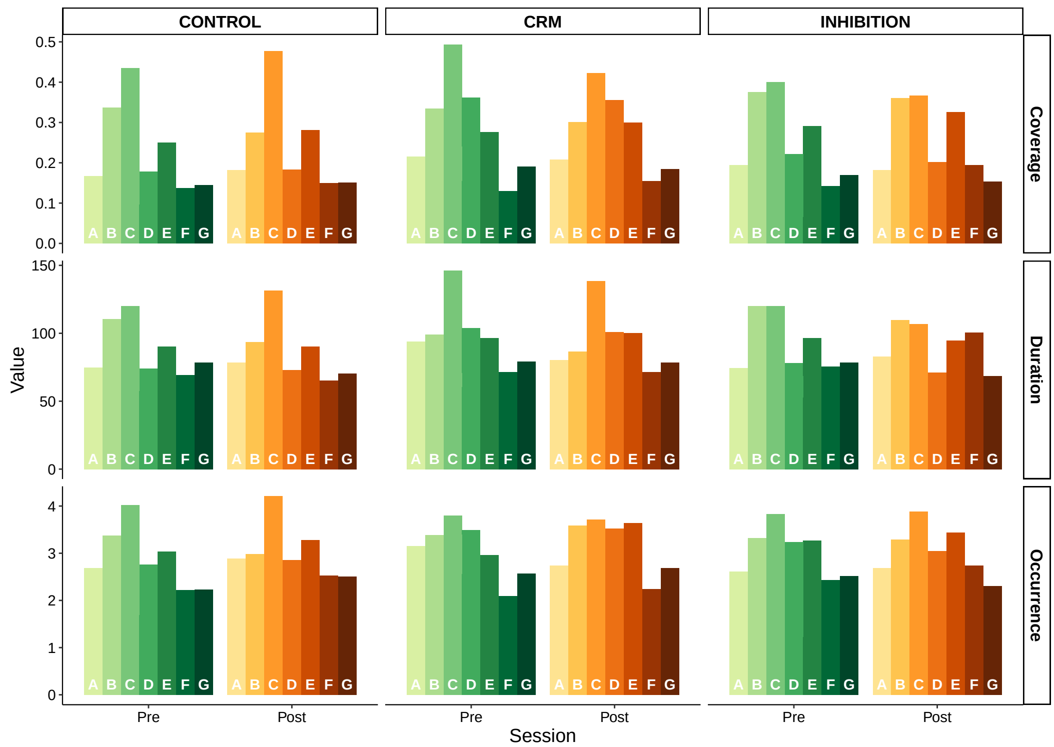


Figure 10.15: Mean coverage (in %), duration (in ms) and occurrence (per second) across all participants for each MS (k=7), for the pre- and post-training sessions

Cognitive Training Microstates ANOVA Table

Coverage

Df: degrees of freedom, F: F-test value, η_p^2: partial eta squared

	DF	F	η_p^2	p
Session	4	0	0	1
Microstates	6	439	.6	$< .001$
Group	2	6	0	1
Session:Microstates	24	.351	.004	.99
Session:Group	8	.17	0	1
Microstates:Group	12	1.07	.01	.381
Session:Microstates:Group	48	.217	.01	1

Table 10.6: Three-way ANOVA for MS (k=7) on Coverage for Session × Group × Microstates

Occurrence

Df: degrees of freedom, F: F-test value, η_p^2: partial eta squared

	DF	F	η_p^2	p
Session	4	.15	.0003	.96
Microstates	6	364.21	.53	$< .001$
Group	2	6	.01	**.003**
Session:Microstates	24	.37	.01	1
Session:Group	8	.17	0	.99
Microstates:Group	12	1.11	.01	.34
Session:Microstates:Group	48	.20	.01	1

Table 10.7: Three-way ANOVA for MS (k=7) on Occurrence for Session × Group × Microstates

Duration

Df: degrees of freedom, F: F-test value, η_p^2: partial eta squared

	DF	F	η_p^2	p
Session	4	.21	.0003	.0004
Microstates	6	199.5	.38	$< .001$
Group	2	9.89	.01	$< .001$
Session:Microstates	24	.24	.003	1
Session:Group	8	.287	.001	.97
Microstates:Group	12	.98	.006	.46
Session:Microstates:Group	48	.004	.01	1

Table 10.8: Three-way ANOVA for MS (k=7) on Duration for Session × Group × Microstates

Investigation and mitigation of listening effort: an electrophysiological and behavioral approach.

Oral communication is at the center of human interaction, and when the auditory scene becomes challenging, listening effort, defined as the ”deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a task” (Pichora-Fuller et al., 2016), varies across individuals. In this project, we investigated listening in effortful auditory environments using both behavioral and electrophysiological approaches. Three experimental phases were conducted: (1) behavioral assessment of speech intelligibility and listening effort in native and non-native speech-in-noise and speech-in-speech conditions (N=51); (2) exploration of the relationship between executive functions and speech listening in challenging conditions, along with investigation of EEG alpha dynamics during effortful listening (N=30); and (3) cognitive training targeting inhibition to improve intelligibility and reduce listening effort, and evaluation of its effects on alpha dynamics (N=60). Behavioral results showed that using a native language enhances intelligibility and reduces listening effort, inhibitory control correlates with performance in the most adverse condition, and cognitive training improves speech perception while decreasing effort. EEG analyses confirmed the involvement of alpha oscillations with diverse neural generators during effortful listening. These findings emphasize the multidimensional nature of listening effort and its critical role in communication. Furthermore, they support the potential of cognitive interventions to mitigate listening challenges, with implications for clinical populations such as individuals with hidden hearing loss.