



HAL
open science

Auditory attention analysis during music listening: making the link between machine learning, electrophysiology and cognition

Joan Belo

► **To cite this version:**

Joan Belo. Auditory attention analysis during music listening: making the link between machine learning, electrophysiology and cognition. Computer science. Université Côte d'Azur, 2022. English. NNT: 2022COAZ4086 . tel-03957453v2

HAL Id: tel-03957453

<https://theses.hal.science/tel-03957453v2>

Submitted on 26 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

Analyse de l'attention auditive lors de l'écoute musicale

Établir le lien entre Machine Learning, Électrophysiologie
et Cognition

Joan BELO

Centre Inria d'Université Côte d'Azur – Équipe-projet Athena

**Présentée en vue de l'obtention
du grade de docteur en** Automatique,
Traitement du Signal et des Images
d'Université Côte d'Azur

Dirigée par : Daniele Schön et Maureen
Clerc

Soutenue le : 05 Décembre 2022

Devant le jury, composé de :

Anne Caclin, Chargée de Recherche, Centre de
Recherche en Neurosciences de Lyon - Rapporteuse
Maureen Clerc, Directrice de Recherche, Centre
Inria d'Université Côte d'Azur - Directrice de Thèse
Rémi Flamarj, Professeur, École Polytechnique –
Examinateur, Président du Jury
Michel Höen, Senior Clinical Director, Oticon Medical
- Examinateur
Diane Lazard, Praticienne Hospitalière, Institut de
l'Audition, Institut Pasteur - Rapporteuse
Daniele Schön, Directeur de Recherche, Institut de
Neurosciences des Systèmes - Directeur de Thèse

PhD Thesis

Auditory attention analysis during music listening

Making the link between machine learning,
electrophysiology and cognition

Joan BELO

Athena Project-Team

Centre Inria d'Université Côte d'Azur

Université Côte d'Azur

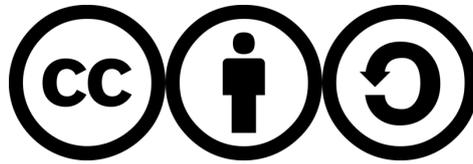
Funding

This doctoral project was funded by Oticon Medical under the research agreement n°14294-CRI04 and The Region Provence-Alpes Côte d'Azur.



License

This document is distributed under a Creative Common License



Attribution-ShareAlike 4.0 International

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material

for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Abstract (Français)

La capacité à analyser des scènes sonores complexes est mise à mal chez les personnes porteuses d'implants cochléaires (IC), notamment dans les environnements bruyants ou en présence de plusieurs sources. Ceci est principalement dû à la faible résolution spectrale de l'IC et au fait qu'il n'amplifie pas uniquement la source d'intérêt.

Une solution possible serait d'utiliser une technique récente, appelée Détection de l'Attention Auditive (AAD), qui permet de détecter, grâce à l'activité neuronale, quelle source, parmi un ensemble de sources, un individu est en train d'écouter. Cette technique s'appuie sur le fait que 1) le cerveau suit l'enveloppe d'amplitude des stimuli auditifs au niveau cortical et que 2) dans les environnements multi-sources, le suivi cortical de la source d'intérêt est amélioré par rapport à celui de la source qui ne l'est pas. Grâce à cette approche, il devient possible d'imaginer une nouvelle génération d'appareils auditifs qui serait capable de détecter la source d'intérêt et de l'amplifier.

Cependant, l'AAD souffre d'une importante variabilité interindividuelle de ses performances de reconstruction et de classification. Cette variabilité peut s'expliquer en partie par des facteurs physiologiques mais aussi potentiellement par des facteurs comportementaux et cognitifs tels, les fonctions exécutives.

Outre le problème de la variabilité, la plupart des études exploitant l'AAD ont utilisé des stimuli de parole. Ici, nous proposons que la musique naturelle polyphonique pourrait permettre de manipuler la complexité de la scène sonore et d'obtenir des informations sur son traitement neuronal.

Nous avons utilisé une approche AAD basée sur la reconstruction de stimulus pour explorer s'il était possible de reconstruire et classifier des extraits musicaux naturels polyphoniques monodiques et dichotiques à partir d'enregistrements EEG. Nous avons aussi mesuré l'attention soutenue, l'inhibition attentionnelle et la mémoire de travail par le biais de tests cognitifs. De plus, nous avons collecté des informations relatives à la familiarité des stimuli ainsi que sur le temps d'errance mentale des participants durant la tâche. Ensuite, nous avons utilisé des modèles de régression linéaire pour voir si les fonctions cognitives, la familiarité et l'errance mentale expliquent les performances du système AAD. De surcroît, nous avons cherché à savoir si l'expertise musicale affecte les

fonctions exécutives et le suivi neuronal des stimuli musicaux en comparant les musiciens et les non musiciens.

Nos résultats ont montré que 1) il était possible de reconstruire et de classifier des stimuli musicaux polyphoniques monodiques et dichotiques, sur la base d'enregistrements EEG, 2) l'inhibition attentionnelle explique environ 10 % des performances dans la condition dichotique, 3) l'expertise musicale ainsi que la mémoire de travail et l'attention soutenue ne sont pas des prédicteurs significatifs de la performance dans les deux conditions, 4) l'expertise musicale n'influence pas les fonctions cognitives et 5) Il existe un effet de la familiarité sur la performance dans la condition monodique et que celui-ci est modulé par le niveau d'errance mentale.

Dans l'ensemble, ces résultats ont montré la faisabilité de l'utilisation de l'approche AAD basée sur la reconstruction de stimulus avec des stimuli musicaux polyphoniques et indiquent que des mécanismes cognitifs spécifiques sont à l'œuvre pendant l'écoute musicale dichotique chez des auditeurs normo-entendants. Ils suggèrent également que l'entraînement de fonctions cognitives particulières pourrait être utile pour améliorer les systèmes AAD et, par extension, les appareils auditifs de prochaine génération.

Mots-clefs : Détection de l'attention auditive, Fonctions Exécutives, Écoute musicale, Science-cognitives, EEG

Abstract (English)

The ability to analyze complex auditory scenes is challenged in cochlear implant (CI) users, especially in noisy environments or in presence of simultaneous auditory sources. This is mainly due to the poor spectral resolution of the CI and to the fact it does not selectively amplify the source of interest.

One possible solution is to capitalize on a recent technique, called Auditory Attention Detection (AAD), that allows to detect, on the basis of neural signals, which auditory source within a set of multiple concurrent sources an individual is attending to. This approach is supported by recent findings showing that 1) the brain tracks the amplitude envelope of auditory stimuli at the cortical level and 2) in multi-sources environments, the cortical tracking of the attended source is enhanced compared to the unattended one. With this approach it becomes possible to imagine a new generation of hearing aids that would be able to detect the source of interest and selectively amplify it.

However, AAD suffers from a high inter-individual variability in the reconstruction and classification performance. This variability can be partly explained by physiological factors but it could also be explained by behavioral and cognitive factors such as executive functions.

Beside the problem of variability, most of the studies exploiting AAD, used only speech stimuli. Here we propose that natural polyphonic music can provide a way to manipulate the complexity of the auditory scene and to gain insights about its neural processing.

We used a reconstruction stimulus-based AAD approach to explore whether it was possible to reconstruct and classify both monodic and dichotic natural polyphonic musical excerpts based on EEG recordings. We also measured sustained attention, attentional inhibition and working memory via cognitive tests. In addition, we gathered information on stimulus familiarity as well as participants' mind wandering during the task. Then, we used linear regression models to see whether cognitive functions, familiarity and mind wandering explain the performance of neural data. Additionally, we explored whether musical expertise affects executive functions as well as neural tracking of musical stimuli by comparing non-musical and musical experts.

Our results show that 1) it was possible to reconstruct and classify ecological polyphonic monodic and, although with lower accuracy, dichotic music stimuli based on single-trial EEG recordings, 2) attentional inhibition explained around 10% of the performance in the dichotic condition, 3) musical expertise as well working memory and sustained attention were not significant predictors of performance in both conditions, 4) musical expertise does not influence cognitive functions and 5) the effect of familiarity on performance was significant in the monodic condition and was modulated by the level of mind wandering of the participants.

Overall, these results showed the feasibility of using the reconstruction stimulus-based AAD approach with ecological polyphonic musical stimuli and point to specific cognitive mechanisms being at work during dichotic musical listening in normal hearing listeners. They further suggest that training specific cognitive functions could be useful to improve AAD systems and by extension, a next-generation of hearing aids.

Keywords: Auditory Attention Detection, Executive Functions, Music listening, Cognitive-science, EEG

Acknowledgment

Il y a eu une multitude d'acteurs qui ont fait que cette thèse a pu voir le jour et surtout, aboutir.

Tout d'abord, je tiens à remercier **Maureen Clerc**. Merci d'avoir eu confiance en moi dès le départ et merci de m'avoir laissé libre de travailler sur des questions qui me plaisaient en dépit de l'orientation clairement différente de ce qui se faisait dans l'équipe Athena.

Ensuite, je remercie chaleureusement **Daniele Schön**. Merci d'avoir accepté de (co)diriger cette thèse (à distance), merci d'avoir tenu la baraque quand c'était la galère, merci d'avoir su me motiver, merci pour tes décisions toujours justes et merci pour tous les conseils prodigués pendant ces trois années. Si c'était à refaire, je le refais juste parce que je sais qu'avec toi, peu importe si c'est dur, on arrivera au bout.

Je remercie tous les collègues de l'équipe Athena, ceux qui sont parti juste après mon arrivée et ceux qui seront encore là quand je serais moi-même parti. Merci à **Abib, Agathe, Amandine, Aymene, Federica, Igor, Ivana, Isa, Johann, Kostiantyn, Ludovic, Matteo, Mauro, Oceane, Patryk, Paul-Emmanuel, Pierre, Petru, Rachid, Sara, Sarah, Sofiane** et **Théo**. Un merci particulier à **Côme** pour les cafés, les séances de stats dans mon bureau et l'aide que tu m'as apportée avec les scripts python et les analyses EEG, à **Samuel** pour ton aide sur les aspects mathématiques du DPI et ta gentillesse et à **Guillaume**, para todas as nossas discussões.

Je remercie aussi les collègues de Marseille qui m'ont aidé de quelque manière que ce soit, pendant cette aventure. Merci à **Benjamin, Céline** et **Jérémy**. Merci à **Christelle** et **Sophie**, pour votre aide sur les triggers de l'EEG. Merci à **Aurélie** pour ton aide sur le manuscrit et les tâches d'attention. Enfin, merci à **Jacques** pour ton aide précieuse sur la TRF.

Merci également à **Eleonora Sulas** et **Dan Gnansia** d'Oticon Medical.

Pour m'avoir permis d'acquérir les données dont j'avais besoin pour ce travail, m'avoir accueilli au Conservatoire Régional de Nice et pour leur sympathie, je remercie **Michel**

Pascal et Gael Navard. Merci également à tous ceux qui ont pris le temps de participer aux expériences.

Je veux aussi remercier tous les poto(e)s, parce que c'est aussi grâce à vous que j'en suis arrivé là. Merci à **Arthur, Charlotte, Clémence, Françou, Loïc, Julie, Juju, Misa, Sarah, Téva et Tomy.** Un merci tous particulier à mon bro **Jim**, toujours là malgré la distance.

Je remercie tout particulièrement **mes parents et mes frères**, pour leur soutien dans les moments difficiles, pour leur amour et pour tous les moments qu'on passe ensemble et qui font du bien. Merci les darons et les bros.

Enfin, je remercie **Emeline**, celle qui me (sup)porte tous les jours, celle sans qui le quotidien n'aurait pas la même saveur, celle sans qui tout ça n'aurait jamais été possible. Merci d'être à mes côtés depuis toutes ces années.

CONTENTS

<i>figures</i>	<i>xvi</i>
<i>Tables</i>	<i>xvi</i>
General Overview	1
General Introduction	3
Sounds and the brain	3
Overview.....	3
What is a sound?	3
The ascending auditory Pathway: from the outer ear to the cortical areas	4
The ear.....	4
The auditory nerve.....	7
The brainstem	7
The Auditory Cortex	10
When the system does not work properly.....	11
The cases of hearing loss	11
Auditory Scene Analysis, Attention and Executive Functions	15
Overview.....	15
Auditory Scene Analysis	15
What is Auditory Scene Analysis?.....	15
Perturbation of auditory streaming.....	18
Bottom-up and top-down factors impacting auditory streaming	19
Attention.....	23
Executive Functions.....	29
Attention, working memory and inhibitory control: A special case?.....	43
Summary	45
Musical Expertise	45
Overview.....	45
Why is musical expertise interesting?.....	45
Influence of musical expertise on executive functions and attention.....	46
Working Memory	47
Inhibitory Control	48
Attention	50
Evidence of the influence of musical expertise on auditory scene analysis.....	51
Summary	53

Auditory Attention Detection	54
Overview.....	54
What is Electroencephalography?.....	54
EEG-based Auditory Attention Detection.....	56
Basics of EEG-based Auditory Attention Detection	56
Two approaches: Linear and nonlinear	56
AAD Performance	59
Limitations.....	60
Rationale behind this work.....	60
<i>Experimental Work</i>.....	63
Study one: Attentional inhibition ability predicts neural representation during challenging auditory streaming.....	63
Introduction	63
Material and Methods.....	66
Participants	66
Main Experimental Task.....	67
Cognitive Tasks.....	69
EEG Data	70
Audio Feature Extraction	71
Stimulus Reconstruction.....	71
Statistical Analysis.....	74
Results	75
Cognitive Functions	76
AAD Performance	77
Effect of Musical Expertise on Reconstruction and Classification Accuracies	82
Effect of Cognitive Functions on Reconstruction and Classification Accuracies	83
Discussion	85
Study Two: The effect of familiarity on neural tracking is modulated by the mind wandering	93
Introduction	93
Material and Methods.....	94
Participants	94
Main Experimental Task.....	95
Behavioral Indicators.....	95
EEG Data	96
Audio Feature Extraction	96
Stimulus Reconstruction.....	96

Statistical Analysis.....	97
Results	97
Behavioral indicators and Performances	97
Discussion	99
General discussion	103
Summary of the results.....	103
On the use of music stimuli	104
On the impact of music stimuli on attentional processes	105
Important excerpts dynamics may have promoted attentional switching.....	105
Absence of bottom-up cues may have favored attentional inhibition	106
On the possible nature and time course of the attentional inhibition mechanism	107
On the importance of working memory and sustained attention for auditory streaming in challenging environments	109
On the role of musical expertise in auditory streaming in challenging auditory situations	110
Perspectives	112
Towards a new generation of cochlear implants?.....	112
Solving a part of the variability problem of AAD performance with cognitive function based predictions.....	113
Improving AAD system's performance with specific cognitive functions-based training.....	114
Is music-based training the best option?.....	116
Other applications of Auditory Attention Detection	117
Conclusion	119
.....	120
Behind the scenes... ..	120
Bibliography.....	125

FIGURES

<i>Figure 1: Description of the human hearing system</i>	4
<i>Figure 2: Cross-section of the cochlea</i>	5
<i>Figure 3: Cross-section of the Organ of Corti</i>	6
<i>Figure 4: Schematic view of the ascending auditory pathway</i>	8
<i>Figure 5: Schematic representation of the auditory cortex and its tonotopic organisation</i>	10
<i>Figure 6: Schematic view of a cochlear implant</i>	13
<i>Figure 7: Illustration of auditory scene analysis</i>	16
<i>Figure 8: Power spectrum density of each musical excerpt</i>	68
<i>Figure 9: Illustration of the effect of musical expertise on cognitive function</i>	77
<i>Figure 10: Heatmaps of reconstruction scores for monodic and dichotic stimuli</i>	79
<i>Figure 11: Illustration of monodic and dichotic reconstruction and classification accuracy</i>	80
<i>Figure 12: Scatterplot of the relation between dichotic and monodic reconstruction</i>	81
<i>Figure 13: Illustration of the effect of musical expertise on monodic and dichotic reconstruction and classification</i>	82
<i>Figure 14: Illustration of the relation between attentional inhibition and dichotic listening metrics</i>	84
<i>Figure 15: Illustration of the main effect of Familiarity on reconstruction accuracy and the simple effects of Familiarity on reconstruction accuracy within the Mind Wandering indicators' level</i>	99
<i>Figure 16: Illustration of the potential timing of the facilitation and inhibition mechanisms</i>	109

TABLES

<i>Table 1: Questions asked to the participants after each music excerpt</i>	96
--	----

GENERAL OVERVIEW

This manuscript is organized in three main chapters. The first chapter is a general introduction where all the notions and concepts required to understand the work done in this thesis are described. It is further divided into five sections.

Section 1 briefly describes the human auditory system and every important relay of the ascending auditory pathway. It also explains what happens when the system does not work properly, describes the different types of hearing loss, their common causes and consequences and makes a zoom on the cochlear implant for which the main advantages and limitations are concisely exposed. Section 2 describes the notion of auditory scene analysis, its principal supportive mechanisms (sequential and simultaneous grouping) and the bottom-up and top-down factors that may influence them. This section also dives into the notion of attention by presenting different views of what attention could be and by providing elements about its key role in the auditory scene analysis. The main models and theories of executive functions, and more particularly working memory and inhibitory control, are also presented here. In addition, it provides information about the multiple ways to measure executive functions as well as the reasons why working memory and inhibitory control are so important for auditory scene analysis. Finally, the end of section 2 focuses on the peculiar relation of working memory, inhibitory control and attention in the context of auditory scene analysis. Section 3 puts emphasis on the controversial impact of musical expertise over executive functions and attention and tries to bring some evidence about why and how musical expertise may impact auditory scene analysis. Section 4 is dedicated to the explanation of the newly developed technique of electroencephalography-based auditory attention detection. In this section, the basic principles of electroencephalography are briefly presented as well as the important discovery about the fundamental properties of the brain that permit the development of auditory attention detection. Afterwards, the different approaches of auditory attention detection are presented (linear and nonlinear) followed by the common performance evaluation criteria and the main limitations. At the end, the last section makes the connection between the previous ones and the next chapter which exposes the work done during this thesis, by outlining the main hypothesis.

The second chapter is dedicated to the experimental works that have been done during the thesis and it is divided into two parts that correspond to two studies.

Finally, the last chapter is a general discussion where a summary of the research questions and the results are first provided. Then, the results are discussed in a more general perspective. Finally, some perspectives for this work are proposed.

GENERAL INTRODUCTION

SOUNDS AND THE BRAIN

OVERVIEW

In this section, we will illustrate the journey of a sound in the human auditory system. Every important step (or relay) of the ascending auditory pathway, starting from the outer ear and ending with the cortical areas of the brain, will be briefly described. Then, we will explain what happens when the system does not work as it should. Therefore, the different types of hearing loss as well as their common causes and consequences will be described. Finally, we will make a zoom on the cochlear implant for which the main advantages and limitations will be exposed.

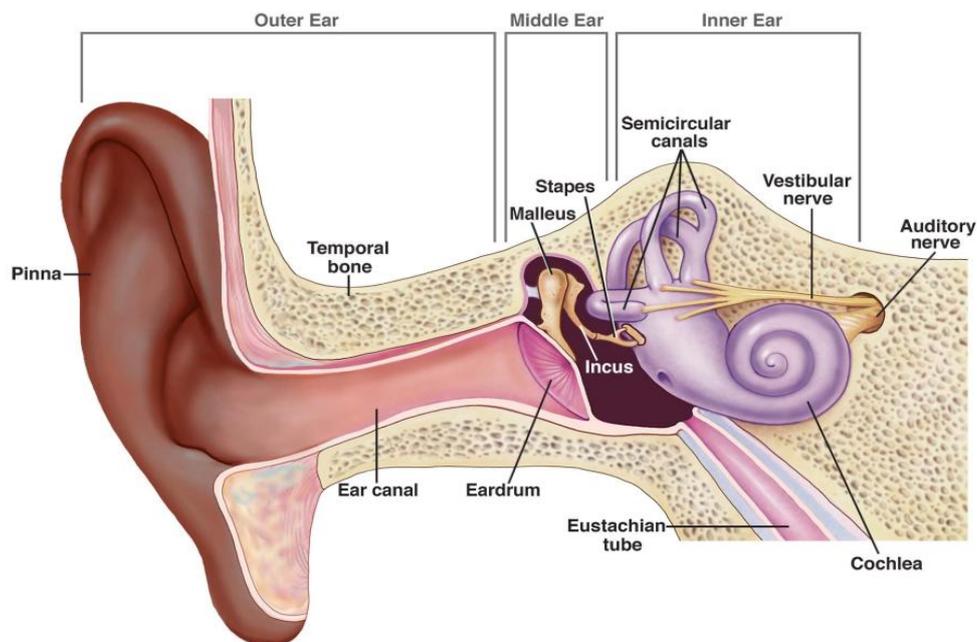
WHAT IS A SOUND?

Physically speaking, a sound is a vibration that propagates through a medium (generally it is air but it can also be a fluid such as water or a solid such as the soundboard of a violin). Because it is a wave, it is characterized by its shape, phase, amplitude and frequency. Without going into details, the shape is commonly expressed as the amplitude as a function of time. The phase gives indication about where the signal is in its cycle. The amplitude corresponds to the intensity of the wave and it is measured in decibels (dB). Finally, the frequency is the number of cycles per second and it is measured in Hertz (Hz). A sound can be more or less complex. In the real-world sounds are generally complex and composed of multiple sinusoidal waves that oscillate at different frequencies. These complex sounds are called harmonic sounds and possess both a fundamental frequency which defines its pitch and a certain number of harmonics which define its timbre. A mathematical technique known as Fourier transform allows one to decompose complex sounds into its sinusoidal components. The human auditory system is able to perceive sound frequency from 20 to 20,000Hz. However, complex sounds are not perceived by human beings as waves but rather as complex percepts. Therefore, how does the auditory system deal with sound waves and transform them into complex perceptions?

THE ASCENDING AUDITORY PATHWAY: FROM THE OUTER EAR TO THE CORTICAL AREAS

THE EAR

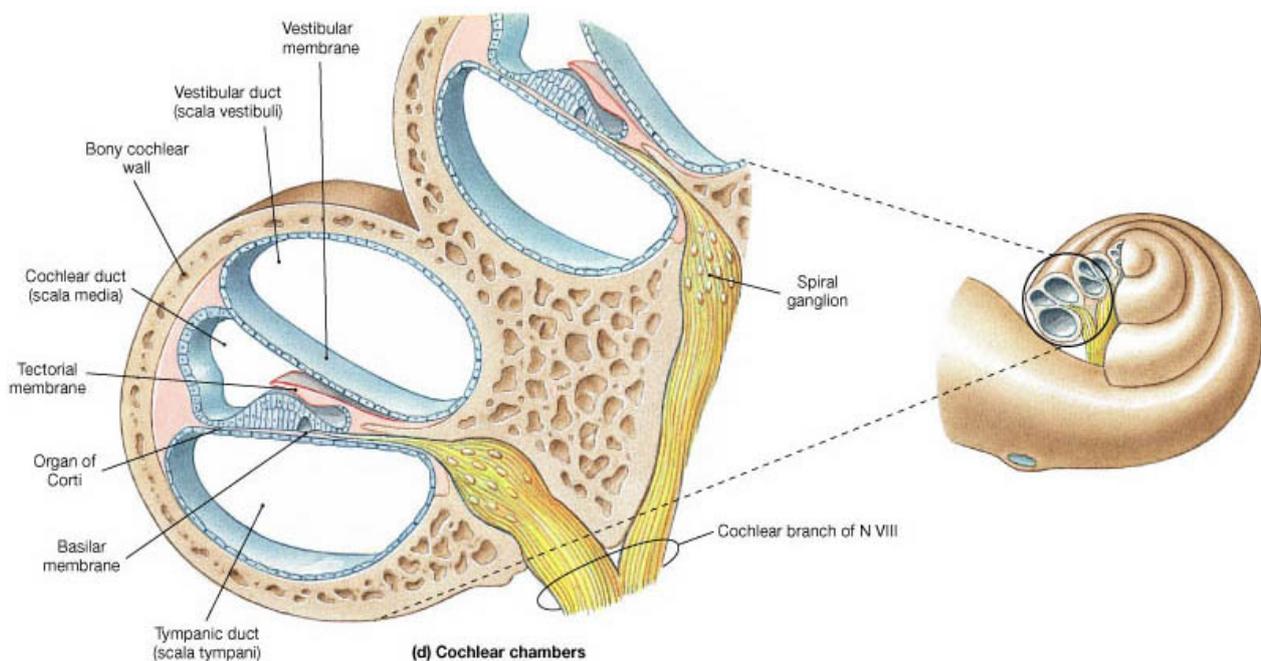
The journey of sounds begins in the external part of the ear, the outer ear, which is composed of the pinna, the auditory canal and the tympanic membrane (Figure 1). The role of the outer ear is to collect sounds and to guide them into the ear canal up to the tympanic membrane which is a very thin membrane (around 70 μ m) that vibrates and transmits sounds to the middle ear.



Source: <https://www.nidcd.nih.gov/health/noise-induced-hearing-loss>

FIGURE 1: DESCRIPTION OF THE HUMAN HEARING SYSTEM DIVIDED IN THREE PARTS: THE OUTER EAR, THE MIDDLE EAR AND THE INNER EAR

The middle ear is the first part of the internal ear. It is mainly composed of a small bones trio called the auditory ossicles, the malleus, the incus and the stapes (Figure 1) and two membranous windows, the oval and the round windows localized on the medial part of the middle ear. The ossicles connect the tympanic membrane to the oval window. The primary role of the ossicular chain is to transfer sound arriving from the tympanic membrane to the inner ear. The ossicles also have a protective function in case of loud noise exposure. In such a case, the ear muscles will contract, reducing the vibration amplitude of the ossicles. The middle ear also comprises the eustachian tube which connects it to the nose and throat. Its function is to maintain normal ambient pressure within the middle ear.

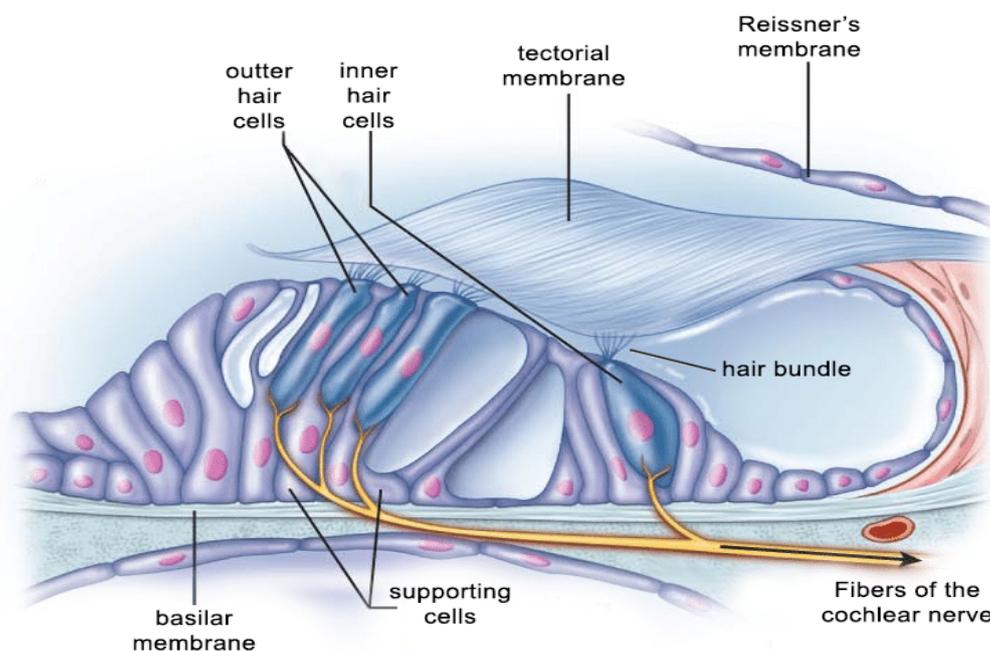


Source: <http://www2.tulane.edu/~h0Ward/BrLg/AuditoryTransduction.html>

FIGURE 2: CROSS-SECTION OF THE COCHLEA

The inner ear is the second and last part of the internal ear. It is composed of the semicircular canal, the vestibule and the cochlea (Figure 1). The semicircular canal and the vestibule form the vestibular system which is the peripheral organ of balance. The cochlea is the focal point of the auditory system (Figure 2). It is a small snail-shaped structure. On the sides of the cochlea are located the scala vestibuli and the scala tympani, two canals filled with fluid, the perilymph, connected to each other by the helicotrema. The scala vestibuli ends with the oval window while the scala tympani ends with the round window. This connection between these two vestibuli allows the perilymph to move back and forth through the helicotrema, when the oval windows oscillate because

of acoustic stimulation. Between the scala vestibuli and the scala tympani is located the scala media. The separation between the scala media and the scala vestibuli is the Reissner's membrane and the separation with the scala tympani is the basilar membrane. The basilar membrane is thicker at the apex and thinner at the base, affecting its vibratory properties (von Békésy, 1960). Consequently, low frequencies will generate vibration at the apex of the basilar membrane while high frequencies will generate vibrations at its base. This is called a tonotopic organization (i.e., a representation of frequencies ordered in space). In addition, a complex sound will generate a specific pattern of vibration similar to the superposition of the vibration generated by all the individual frequencies of which it is constituted. Inside the scala media are also located the tectorial membrane as well as the organ of Corti (Figure 3).



Source : Dominguez-Morales., 2018

FIGURE 3: CROSS-SECTION OF THE ORGAN OF CORTI

The function of the organ of Corti is to transform the vibration of the basilar membrane to an electric signal which will be sent to the auditory nerve. It is roughly composed of sensory and supporting cells. There are two types of sensory cells, called hair cells: the outer hair cells and the inner ones. There are around 3500 inner hair cells compared with 12000 outer hair cells. In reality, the true sensory receptors are the inner hair cells that

receive 95% of the afferent auditory nerve fibers. It is the hair cells that transmit the electric signal to the auditory nerve.

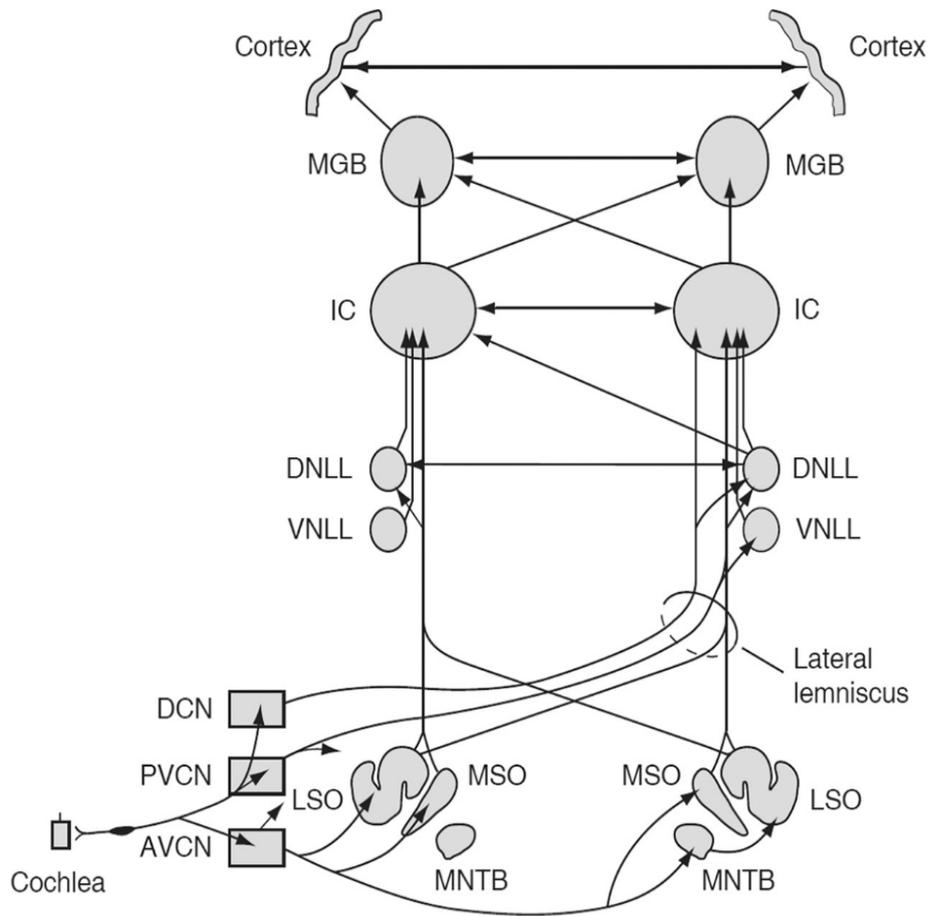
THE AUDITORY NERVE

The auditory nerve (or cochlear nerve) connects the brain to the organ of hearing and balance. The auditory nerve is directly connected to the hair cells via the ascending neurons of the spiral ganglion. Like the basilar membrane of the cochlea, the auditory nerve follows a tonotopic organization. Fibers coming from the apex of the cochlea (i.e. sensitive to low frequencies) are located in the center while those coming from the cochlea's base (i.e. sensitive to high frequencies) are located at the periphery of the auditory nerve (Helfert et al., 1991). The auditory nerve merges with the vestibular nerve to form the vestibulocochlear nerve and reaches the cochlear nucleus in the brainstem.

THE BRAINSTEM

The cochlear nucleus represents the first relay of the ascending auditory pathway in the brainstem (Figure 4).

Interestingly, it is also tonotopically organized. The anteroventral part of the cochlear nucleus gives rise to the ventral auditory stream of the brainstem, which is involved in the localization of binaural sounds while the dorsal part of the cochlear nucleus gives rise to the dorsal auditory stream of the brainstem, engaged in the analysis of complex stimuli.



Source: Pickles, 2015

FIGURE 4: SCHEMATIC VIEW OF THE ASCENDING AUDITORY PATHWAY FROM THE FIRST RELAY (COCHLEAR NUCLEUS) TO THE CORTEX.

The next relay of the ascending auditory pathway in the brainstem is the *superior olivary complex* (Figure 4). It is a part of the ventral auditory stream which receives bilateral innervation allowing for binaural processes (Amunts et al., 2012). For instance, the *superior olivary complex* is involved in sound's lateralization by comparing the intensity and the difference in timing of the stimuli between the two ears (Pickles, 2015; Golding and Oertel, 2012). It is also involved in the processing of high (Caspary and Finlayson, 1991; Cant and Hyson, 1992; Glendenning and Masterton, 1998) and low frequency (Goldberg and Brown, 1968; Warr, 1982; Yin and Chan, 1990; Smith et al., 1993; Spitzer and Semple, 1995). Importantly, the dorsal auditory stream does not pass through the superior olivary complex. Instead, it takes another road. However, it is important to mention its involvement in extracting spectral characteristics of sounds and localizing sounds in the vertical plane (Sutherland et al., 1998).

The inferior colliculus is the largest auditory nucleus of the human auditory system (Amunt et al., 2012) and the third relay of the auditory pathway. While it is supposedly involved in the detection of novel sounds (Lumani and Zhang, 2010), it is also an important structure for acoustic–motor coordination (Olazabel and Moore, 1984) and somatosensory integration (Jain and Shore, 2006). Importantly, it is the first location, in the auditory pathway, where ventral and dorsal auditory streams converge. Therefore, the inferior colliculus is the fusion site of the acoustic stimulus properties carried by the two streams, i.e., localization of sounds properties for the ventral stream and identification of sounds for the dorsal stream.

Finally, The Medial Geniculate Body is the last relay before the cortex. It can be subdivided into two main nuclei, the principal and the medial nucleus (Jones, 2003). The medial geniculate body role is to refine the integration of all the acoustic stimulus' features (spatial, spectral and temporal) that have been started in the *inferior colliculus*. It represents a step forward in the creation of an auditory object. For its part, the medial nucleus is involved in multisensory integration and receives connections from a variety of sources such as the central nucleus of the inferior colliculus, the lateral tegmental system or the vestibular system. At the end, the acoustic stimulus leaves the brainstem by the medial geniculate body to reach the ipsilateral temporal gyrus.

THE AUDITORY CORTEX

The auditory cortex (Figure 5) is the main target of ascending fibers that are coming from the medial geniculate body and the final link of the ascending auditory pathway. It is located on the temporal lobe and encompasses two-thirds of the superior temporal gyrus. It can be decomposed into the primary auditory cortex (A1) which occupies the major part of the Heschl gyrus and the secondary auditory cortex (higher order auditory areas) which refers to a belt of peripheral auditory areas surrounding the primary auditory cortex.

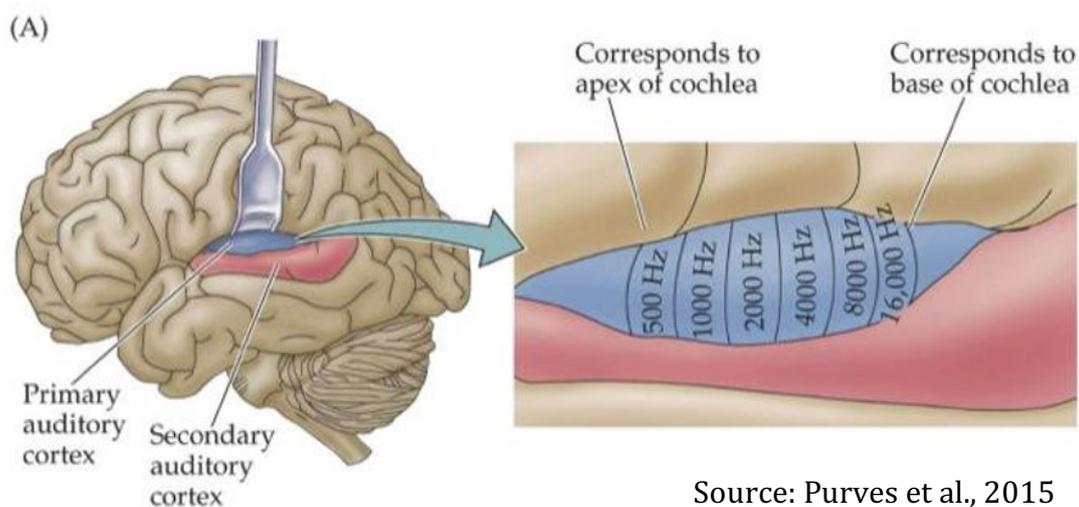


FIGURE 5: SCHEMATIC REPRESENTATION OF THE AUDITORY CORTEX AND ITS TONOTOPIC ORGANISATION

The primary auditory cortex is tonotopically organized (Formisano et al., 2003; Talavage et al., 2004; Woods et al., 2009, 2010; Humphries et al., 2010; Da Costa et al., 2011; Striem-Amit et al., 2011; Langers and van Dijk, 2012) and receives projections of the ventral part of the principal nucleus of the medial geniculate body while the areas of the belt mainly receive projections from its dorsal part. Nevertheless, the precise organization as well as function of the human auditory cortex is not yet completely clear and there is no consensus in the literature about both the anatomical-based subdivision and the functional-based one (Moerel et al., 2014; Hackett, 2015). For instance, it would appear that the left and right auditory cortex are structurally (Penhune et al., 1996; Hutsler and Gazzaniga, 1996; Anderson et al., 1999; Galuske et al., 2000) and functionally asymmetric (Cammoun et al., 2015; Zatorre, 1988; Belin et al., 2000; Hickok and Poeppel, 2000; Scott

et al., 2000) even if this asymmetry is not fully understood yet. For example, music and speech could partly be processes in opposite hemispheres (Zatorre et al., 2002; Poeppel, 2003, Albouy et al., 2020). This could arise from a “sensitivity” of each hemisphere (Morereel et al., 2014, Albouy et al., 2020), the left hemisphere being rather sensitive to temporal processing (Shannon et al., 1995; Liégeois-Chauvel et al., 1999; Zatorre et al., 2002) and therefore in speech processing while the right hemisphere could be more sensitive to spectral processing (Liégeois-Chauvel et al., 2001; Zatorre and Belin, 2001; Zatorre et al., 2002) and consequently in music processing.

WHEN THE SYSTEM DOES NOT WORK PROPERLY

THE CASES OF HEARING LOSS

Until now, we describe the functioning of the normal auditory system, meaning when a sound is able to go from the outer ear to the auditory cortex. However, sometimes, it does not work properly. The World Health Organization estimates that around 1,5 billion people suffer from hearing loss with around 430 million having disabling hearing loss. Therefore, it is a real societal problem. The reasons why the auditory system can falter are diverse. And the consequences of these malfunctions can also be very different. However, one can distinguish three main types of hearing loss, conductive, sensorineural and mixed hearing loss (Zahnert, 2011).

CONDUCTIVE HEARING LOSS

Conductive hearing loss is characterized by the impossibility for an acoustic stimulus to propagate from the outer ear to the inner ear. Therefore, it concerns both the outer and middle ear. The reasons can be genetic or acquired. Congenital conductive hearing loss can be for example the result of incomplete formation of the outer ear due to aural atresia (Abdel-Aziz, 2014) or due to otosclerosis which causes a calcification of the stapes. However, conductive hearing loss is generally acquired. The main cause is the Otitis media with effusion (Coleman and Cervin, 2019) but it can also be due to the presence of an obstacle in the ear canal like wax or foreign bodies or due to a perforation of the tympanic membrane (Sogebi et al., 2017).

Conductive hearing loss is generally transient but, in some instances, it requires surgery. It is the case in otosclerosis, when the ossicles need to be removed and replaced by a prosthesis via a stapedectomy (Quesnel et al., 2018). When there is no medical or surgical

treatment, conductive hearing loss can be treated with bone conduction hearing aids such as bone-anchored hearing aids (BAHA) which transmit sounds directly to the cochlea via vibrations of the skull bone.

SENSORINEURAL HEARING LOSS

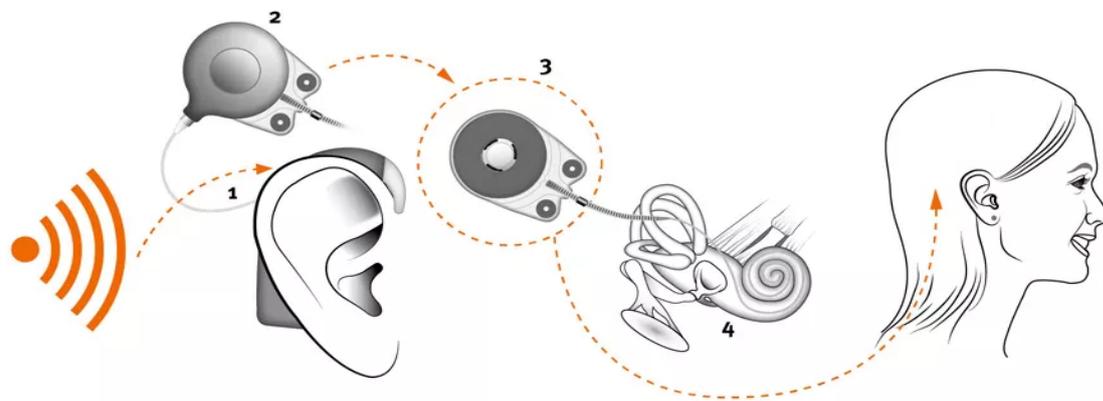
Sensorineural hearing loss refers to “any cause of hearing loss due to a pathology of the cochlea, auditory nerve, or central nervous system” (Tanna et al., 2022). Sensorineural hearing loss can also be congenital or acquired. In congenital cases, the main reasons are genetic (Marazita, 1993) such as GJB2 gene mutation (Snoeckx et al., 2005) or syndromes like Down or Usher's syndrome. But other factors such as congenital infections (e.g., cytomegalovirus), drug abuse, malformations of the inner ear, can lead to hearing loss in fetuses. When acquired, the reasons can be diverse. In children, infections (e.g., rubella, cytomegalovirus, toxoplasmosis) occurring near after birth are the most prevalent reasons (Kenna, 2015). In adults, besides head trauma that can disrupt the structure of the inner ear, common causes are exposure to loud noise, which cause the destruction of hair cells (Carroll et al., 2017), ototoxic drugs consumption and aging (known as presbycusis). Presbycusis is characterized by a bilateral progressive deterioration of auditory sensitivity, mainly in the high frequency and it affects around 23% of the population between 65 and 75 years and 40% older than 75 years (Seidman et al., 2002).

Sensorineural hearing loss are generally permanent and therefore cannot be treated with medication. However, depending on the severity of the hearing loss, solutions exist. For mild to middle hearing loss, and when the structures of the auditory system are undisturbed, the most common solution is the conventional hearing aid. These are sound-amplifying devices that increase the user ability to perceive sounds in the environment. The most famous example is the “behind the ear” hearing aid. For severe to profound hearing loss and when the cochlea is not functioning properly anymore the solution is the cochlear implant.

ZOOM ON THE COCHLEAR IMPLANT

The history of the cochlear implant began in the second half of the 20th century with the first patient implantations in the 70's. Nowadays, around 750 000 people have been successfully implanted with such a device. The cochlear implant is a device created to substitute the peripheral auditory system (from the outer to the inner ear) when it is not

working properly. It is composed of several elements (figure 6) some of them being external and others internal. The first piece that takes place behind the ear is a microphone (1) which captures, digitalizes and encodes the ambient sounds as radio frequencies. The microphone is connected to an external antenna (2) which will send the encoded sound to the internal part of the implant (3) located under the skin behind the ear. This internal part is a stimulator that interprets the radio frequencies sent by the antenna, transforms them into electric currents and sends them into an array of 12 to 22 electrodes (4) directly implanted along the cochlea which will, in turn, stimulate the auditory nerve, “replacing” the defective cochlea.



Source : Oticon Medical

FIGURE 6: SCHEMATIC VIEW OF A COCHLEAR IMPLANT. 1) SOUND PROCESSOR. 2) ANTENA, 3) UNDER THE SKIN MAGNETIC PART OF THE IMPLANT, 4) ELECTRODE ARRAY.

Although cochlear implants are efficient and allow users to perceive sounds, they suffer from several limitations. For instance, they suffer from distortions in the frequency domains (Macherey and Carlyon, 2014). In fact, the electrodes implanted in the cochlea generally don't stimulate independent neural populations because of many reasons (e.g., fluid-filled environment, cross-turn stimulation). Consequently, spatial specificity in the frequency domain is quite poor and cochlear implant users have difficulties in differentiating close frequencies. In addition, electrode arrays are generally implanted along the first one-and-a-half turns of the cochlea, leading to a lack of stimulation of its most apical part and therefore representation of lower frequencies are not provided. Cochlear implants have also some caveats in the temporal and amplitude domains mainly

due to the pulse strategies adopted to stimulate the nerve fibers and the properties of their membrane as well as of these fibers themselves (Macherey and Carlyon, 2014; Lorenzi et al., 2006). These technical limitations engender difficulties in noisy conditions. While in quiet situations, several studies demonstrated that speech comprehension benefits from the implantation of cochlear implants (see for example Leigh et al., 2016 and Holden et al., 2013), in noisy environment, sound localization (Kerber and Seeber, 2012; Williges et al., 2019) and speech comprehension (Caldwell and Nittrouer, 2013; Schleich et al. 2004; Loizou et al. 2009; Misurelli and Litovsky, 2015; Wilson, 2018; Nelson et al., 2003; Sladen and Zappler, 2015; Nelson and Jin, 2004) remain difficult. This is mainly because speech comprehension in very complex acoustic environments is based on the separation of the target sound from the background which rely upon spectral, spatial and temporal cues that are difficult to use by cochlear implant users. In addition, cochlear implants do not allow selective amplification of the source of interest, exacerbating the poor auditory perception in noisy environments or in presence of simultaneous auditory sources (Moore, 2003; Crandell, 1991; Humes et al., 1996).

Another situation where cochlear implant users are prejudiced is music listening because music comprehension requires good pitch perception (McDermott, 2004; Limb and Roy, 2013). While several attempts have been made to make music more accessible for cochlear implant users using for instance, pitch reduction techniques (Nagathil et al., 2017), vocal boosting (Buyens et al., 2014; Pons et al., 2016; Gajęcki and Nogueira, 2018) or even music listening training (Driscoll, 2012; Fu and Galvin, 2007; Gfeller et al., 2002; van Besouw et al., 2014), music perception remains quite poor after cochlear implantation (Prevoteau et al., 2018) and the inter-individual variability in music enjoyment is very important (Kohlberg et al., 2014; Migirov et al., 2009; Mirza et al., 2003).

AUDITORY SCENE ANALYSIS, ATTENTION AND EXECUTIVE FUNCTIONS

OVERVIEW

In this section we will discuss three central concepts, auditory scene analysis (ASA), Attention and Executive functions (EF). First, we will give a definition of ASA as well as the main processes underlying it. Then, we will talk about the principal bottom-up and top-down factors that influence these processes with an emphasis on attention. Afterwards, we will focus on executive functions, propose a definition and go into more details about what working memory and inhibitory control are, how we can measure them behaviorally and why they are important for auditory streaming. Finally, we will explain the reasons why working memory, inhibitory control and attention have a very special relationship.

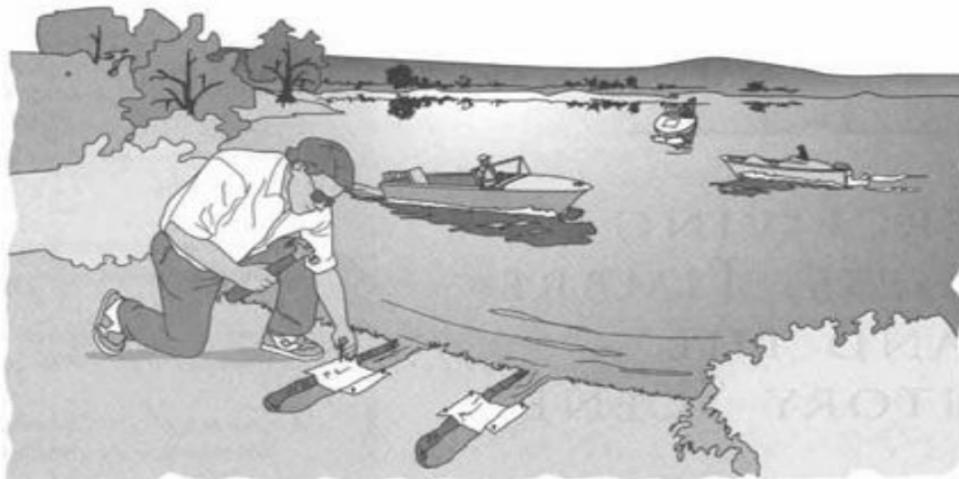
AUDITORY SCENE ANALYSIS

WHAT IS AUDITORY SCENE ANALYSIS?

In real life, it is very frequent to encounter situations where multiple acoustic sources are present at the same time. It is for example the case for someone having a conversation with a friend in a bar, with some music playing in the background and other conversations taking place concurrently. In this kind of situation all the sounds are mixed together. Therefore, it is this mixture that reaches the ears of the listener and not the individual sounds composing the acoustic environment. This is what Cherry (1953) named the “cocktail party problem” in the early 50’s. However, and rather hopefully, human beings are able to perceive the individual sound sources composing this incoming mixture. This is what was called Auditory Scene Analysis and defined by Bregman (1990) as “*the process by which the auditory system separates the individual sounds in natural-world situations, in which these sounds are usually interleaved and overlapped in time and their components interleaved and overlapped in frequency*”. Since the seminal book of Bregman in 1994, ASA has become a conceptual framework that encompasses numerous notions and disciplines and allows multidisciplinary research (Alain and Bernstein, 2015). However, the vocabulary is still not unified and one can find in the literature several words that are used interchangeably, to discuss the same concept. It is for example the case of terms *grouping* and *integration* even if sometimes the first one refers to the

process whereas the other one is the result of this process. Therefore, when necessary, we will provide precision about the terms employed in this manuscript via footnotes.

One of the most critical functions to analyze auditory scenes is the ability to decompose an acoustic mixture into auditory events (for instance a note or a syllable), then to group them into *streams*, which are sequences of auditory events that are perceived as coherent entities (for instance a melody or a speech), and finally to follow these streams through time. With the assumption that each stream is emanating from a different sound source. This process is named *auditory streaming* (figure 7). The formation of distinct auditory streams is underlied by two mechanisms, *simultaneous grouping* and *sequential grouping* (Bregman, 1990).



Source: <https://www.cns.nyu.edu/~david/courses/perception/lecturenotes/localization/localization.html>

FIGURE 7: ILLUSTRATION OF AUDITORY SCENE ANALYSIS. THE LAKE CORRESPONDS TO YOUR AUDITORY WORLD, THE WAVES ON THE LAKE CORRESPOND TO SOUND WAVES, THE TWO CHANNELS CORRESPOND TO YOUR EAR CANALS, AND THE TWO PIECES OF CLOTH CORRESPOND TO YOUR TWO EARDRUMS. JUST FROM THE MOTION OF THE CLOTHES, YOU HAVE TO FIGURE OUT WHAT'S HAPPENING ON THE LAKE. DAVID HEEGER, 2006.

SIMULTANEOUS GROUPING

Simultaneous grouping is the mechanism by which sound elements that overlap in time are grouped together to form an auditory event (Ciocca, 2008). Simultaneous grouping is necessary because natural sounds are generally composed of several frequencies.

Therefore, the auditory system has to group (or integrate¹) frequencies arising from the same source to be able to segregate auditory events that are coming from different sources. This mechanism operates at a local scale (Shinn-Cunningham et al., 2017).

SEQUENTIAL GROUPING

Sequential grouping, on the other hand, is the mechanism by which auditory events are grouped in sequence to form an auditory stream (Ciocca, 2008) and it operates at a longer time scale (Shinn-Cunningham et al., 2017)². This mechanism is related to the fact that natural sounds are dynamic, meaning that they are evolving in time. Consequently, it is important for the auditory system to group (or integrate) sounds in time, otherwise, human beings would not be capable of following a conversation.

A good example of the phenomenon of *auditory streaming* was demonstrated using the now-famous ABAB... or ABA-ABA-... paradigm³ (Miller and Heise, 1950; Bregman and Campbell, 1971; van Noorden, 1975). In this experimental paradigm, alternating pure tones of different frequencies (e.g., a tone A at 200Hz and tone B at 300Hz) are either perceived as a single or two streams depending on the perceptual distance between the two tones (Bregman, 1990). When this perceptual distance is low, for instance when the frequency difference between A and B is small (e.g., a semitone), the alternating sequence will be perceived as a single stream. This is known as *stream integration* or *fusion* (Moore and Gockel, 2012). However, when the frequency difference between A and B is larger (e.g. an octave), the sequence of tones will be perceived as two different streams. This is the phenomenon of *stream segregation* or *fission* (Moore and Gockel, 2012). In the intermediate situation, bistable perception will occur and the sequence will be alternatively perceived as one or two streams (Pressnitzer and Hupé, 2006). The

¹ Note that we will use the term *grouping* and *integration* interchangeably in the rest of this manuscript.

² Note that simultaneous and sequential grouping are probably not independent (Bregman, 1990). However, in this manuscript we choose to separate them both for convenience and because their interdependency is not completely understood.

³ “-” corresponds to a silence.

perceptual distance is not solely defined by the frequency difference between the tones but it can also result from a combination of other properties such as the repetition rate of the tones, inter-tone interval time, or their number in the sequence (Micheyl and Oxenham, 2010). Since its first demonstration, auditory streaming have been documented for other types of sounds such as harmonic complex tones (e.g., Singh, 1987; Cusack and Roberts, 1999; Vliegen et al., 1999; Vliegen and Oxenham, 1999; Grimault et al., 2000, 2001; Roberts et al., 2002, 2008; Bregman and Levitan, 1983), music (e.g., Marozeau et al., 2013; Wright and Bregman, 1987; Rasch, 1978; Uhlig et al., 2013), synthetic vowels (e.g., Gaudrain et al., 2007, 2008), speech (e.g., Dupoux et al., 2008; Darwin et al., 2003; Lee and Humes, 2012; Gordon-Salant and Fitzgibbons, 2004) and even noise (e.g., Bregman et al., 2001; Grimault et al., 2002).

PERTURBATION OF AUDITORY STREAMING

In real-life settings, auditory streaming can be disturbed by several phenomena. This is for example the case where “energetic masking” is present in the auditory scene (Shinn-Cunningham et al., 2017). It refers to a situation where the target and distractor(s) overlap in the spectro-temporal domain (Brungart, 2001). They are, consequently, in competition in the peripheral auditory structures (i.e., cochlea and auditory nerve) because both target and distractor(s) elicit the same, or overlapping, regions (*peripheral channeling hypothesis* from Hartmann and Johnson, 1991).

It is also the case in the presence of “informational masking”. Informational masking could arise in situations where several competing salient signals are present in the auditory scene (Shinn-Cunningham, 2008).

Finally, targets and distractors can also emerge from the same location, making it more difficult to disentangle them (Middlebrooks, 2017).

Certain ecological contexts are more or less prone to generate integration of streams. This is particularly the case in multi-part music listening where segregation of the composing streams is made more difficult by their similarity in harmony, timbre (Pressnitzer et al., 2011; Ragert et al., 2014) or spatial location. That being said, multi-part music is also different from speech in noise listening for instance, because it allows the listener to choose between listening to a specific instrument or listening to the whole piece which is achieved by integrating the different melodic lines (Gregory, 1990; Bigand et al., 2000).

This makes polyphonic music a good testing ground to explore integration across complex sounds (Uhlig et al., 2013; Ragert et al., 2014; Disbergen et al., 2018).

BOTTOM-UP AND TOP-DOWN FACTORS IMPACTING AUDITORY STREAMING

Yet, a big question remains unanswered: what are the main factors that make auditory streaming possible in complex acoustic environments such as speech in noise or music listening?

In fact, there are two sorts of factors that influence auditory streaming, *bottom-up*, also named *low-level*, *primitive* or even *pre-attentive* factors and *top-down*, also named *high-level* or *schema-based* factors⁴ (Bregman, 1990; Moore and Gockel, 2012). To make this more complicated, bottom-up and top-down factors can interact with each other, especially in real-life situations where a multitude of overlapping complex sounds are composing the acoustic environment. However, this interdependency is still not fully appreciated.

BOTTOM-UP FACTORS

The idea behind bottom-up factors is that the auditory system will automatically use some basic characteristics of the different sound sources present in the acoustic environment to either segregate or integrate them. In reality, bottom-up factors can operate either at the level of the auditory events and/or at the level of the auditory streams. But no matter at what level they operate, they always obey a “spectro-temporal proximity law”. This proximity law is quite analogous to the *similarity principle* of the Gestalt psychology in the visual modality (Köhler, 1947) which says that similar elements (such as shape or colors) tend to be grouped together into clusters. Therefore, sound elements with similar characteristics in frequency and time will tend to be automatically integrated together in a unique auditory event and auditory events with similar characteristics will be integrated in a unique auditory stream. In contrast, sound elements with different characteristics in frequency and time will tend to be segregated into

⁴ Note that we will use the term *bottom-up factors* and *top-down factors* in the rest of this manuscript.

separate auditory events and auditory events with different characteristics will be segregated into separate auditory streams (Bregman, 1990).

At the auditory event level (or simultaneous grouping level), several basic features of sounds can force their integration or segregation into unique or different auditory events (see Ciocca, 2008 for a complete review). For instance, previous studies using Event-Related Potential (ERP) and Auditory Evoked Fields (AEF) measurements found that harmonicity impacts the formation of auditory events. Harmonic sounds were more often perceived as unique auditory events than inharmonic ones (Alain et al., 2001; Alain and McDonald, 2007; Bidet-Caulet et al., 2007; Hartmann et al., 1990; Moore et al., 1986). Onset synchrony, or “common onset”, is another factor that promotes integration of sounds into the same auditory event (Bregman, 1990; Ciocca, 2008; Lipp et al., 2010). This is one of the most influential factors because the different parts (i.e., the sounds) of a single auditory event (e.g., a note) generally start at the same time, more or less a few milliseconds (Bregman, 1990). In contrast, offset synchrony (or “common offset”) has been shown to have a marginal impact on the integration of frequency components into auditory events (Dannenbring and Bregman, 1978; Darwin, 1984). Another factor that impacts auditory event formation is the spatial location of the sounds. Indeed, sounds emanating from the same location tend to be grouped together in a single auditory event (Bregman, 1990). However, it seems not to be a crucial cue for sounds integration when other cues are available (Smith et al., 1982; Darwin and Hukin, 1997; Shinn-Cunningham et al., 2007; Schwartz et al., 2012; Shinn-Cunningham et al., 2017). Finally, because sounds arising from the same source will have a common amplitude modulation, coherent amplitude fluctuation is also an important factor that helps the auditory system to integrate or segregate sounds into auditory events (Bregman, 1990; Ciocca, 2008; Shinn-Cunningham, 2017; Sheft, 2007).

At the auditory stream level (or sequential grouping level), multiple characteristics of auditory events can also promote their integration or segregation into streams. The first one is the difference between the fundamental frequencies (Δf_0) of the auditory events (the Δf_0 corresponds to the pitches). Van Noorden (1975), using the classical ABA-ABA... ABAB... paradigm, found that if the frequency separation between tones is larger than a certain threshold, then tones will be perceived as two separate streams but if the frequency separation is lower than a certain threshold, tones will be perceived as a

unique stream. He named these thresholds the *temporal coherence boundary* and the *fission boundary*. More recently, several works have shown that the larger the Δf_0 , the easier the segregation of concurrent speech streams (e.g., Alain et al., 2005; Chalikia and Bregman, 1989; Culling and Darwin, 1993). And in music, the influence of pitch on segregation of streams was also demonstrated (e.g., Deike et al., 2004; Dowling, 1973; Hartmann and Johnson, 1991; Marozeau, et al., 2010). Spectral content (timbre) is also a powerful cue that allows segregation or integration of auditory events. For instance, Cusack and Roberts (2000) found that timbre differences between sounds in a sequence have an impact on their segregation into separate streams. This result is in line with previous studies (e.g., Singh, 1987; Bregman et al., 1990; Iverson, 1995; Singh and Bregman, 1997). Impact of timbre on the formation of auditory streams was also illustrated with music (e.g., McAdams, 2003; Marozeau et al., 2013; McAdams, 2013) as well as with speech (e.g., Lackner and Goldstein, 1974; Culling and Darwin, 1993). As for auditory event formation, temporal envelope differences (specifically onset timing and amplitude modulation) are also prominent factors for stream constitution (Grimault et al., 2002; Carlyon, 2004; and see Moore and Gockel, 2012 for a complete review). Interestingly, spatial location, while not crucial in the auditory events formation, is important in the integration and segregation of those auditory events into streams (Darwin, 2006; Maddox and Shinn-Cunningham, 2012). It was first coined by Cherry in 1953 and many studies since have demonstrated the importance of spatial location. For example, Hartmann and Johnson (1991) provided evidence that a $\pm 500\mu\text{s}$ Interaural Timing Difference (ITD) is sufficient to segregate two melodies (see also Sach and Bailey, 2004; Saupe et al., 2010; Boehnke and Phillips, 2005 and Darwin and Hukin, 1999). Moreover, spatial cues are of utmost importance in speech and speech-in-noise contexts (e.g., Ihlefeld and Shinn-Cunningham, 2008; Kidd et al., 2008; Middlebrooks, 2017; Hawley et al., 2004; Plomp and Mimpen, 1981).

Finally, a quite recent theory, the *temporal coherence theory*, proposed by Shamma and colleagues (2011), claims that stream segregation is not solely based on separation in feature-space (e.g., location, pitch, timbre) but rather on the temporal relationship between the different auditory objects in the scene. In their model, the formation of auditory streams depends upon the *temporal coherence* of responses of neural populations that are sensitive to various features of a sound (e.g., frequency, pitch, timbre,

location). This means that temporally coherent auditory objects are grouped together while temporally incoherent ones are separated and that this temporal coherence analysis operates at a neural level. This theory has been substantiated by psychoacoustic, neuroimaging as well as computational studies (e.g., Elhilali et al., 2009a; O’Sullivan et al., 2015; Lu et al., 2017; Christiansen et al., 2014, Teki et al., 2013), such as the princeps work of Elhilali and colleagues (2009b) who found that two sequences of synchronous tones were not segregated even when their frequency separation was large. Interestingly, although temporal coherence analysis seems to be an automatic neural process, the authors proposed that top-down factors, and more specifically attention, can modulate it.

TOP-DOWN FACTORS

In contrast to bottom-up factors, it is way less clear whether top-down factors influence auditory stream formation or if they come into play once the streams have been formed. However, to provide some details about this ongoing debate it is required to distinguish two types of top-down factors. In fact, in the taxonomy proposed by Bregman (1990), top-down factors encompass simultaneously processes related to “conscious attention” and those related to prior knowledge about sounds. That being said, the main debate about the influence of top-down factors on auditory streaming concerns attention more than prior knowledge (Shamma and Micheyl, 2010; Shamma et al., 2011; Snyder et al., 2012). Therefore, in the next sections we will separately discuss the impact of prior knowledge and attention on auditory streaming.

PRIOR KNOWLEDGE

The possible impact of prior knowledge⁵ on auditory streaming has interested researchers since the early 70’s and the work of Dowling (1973) who found that when two melodies were close in pitch, more familiar melodies were easily segregable from other melodies than non-familiar ones. Since this foundational work, several studies have provided insight about the influence of prior knowledge on auditory streaming. For instance, when the target talker is familiar, speech perception is improved in the presence of concurrent talkers (e.g., Johnsrude et al., 2013; Souza et al., 2013; Newman and Evers,

⁵ Note that we will use the terms *prior knowledge* and *familiarity* interchangeably in the rest of the manuscript.

2007, Barker and Newman, 2004, Yonan and Sommers, 2000; Wang et al., 2019). And it is also the case for the comprehension of speech in noise (e.g., Nygaard and Pisoni, 1998; Nygaard et al., 1994). This influence of prior knowledge over auditory stream formation was also documented for music stimuli. Indeed, Bey and McAdams (2002) have shown that the ability to segregate interleaved melodies was impacted by the knowledge of the melody. When the participants had precise knowledge of melody to recognize within the mixture, then segregation was better compared to when participants only had general knowledge (i.e., information about the pitch of the melody). A very recent study also demonstrated, using EEG measurement, that cortical representation of familiar songs was better than cortical representation of unfamiliar ones (Vanden Bosch der Nederlanden et al., 2022). Nevertheless, this has to be shadowed by the fact that other studies, using the same approach, found opposite results (Kumagai et al., 2017, 2018).

In view of these elements, it seems that prior knowledge of the stimuli can influence auditory streaming. That being said, even if some insights are available in the literature to say that “short-term knowledge” could be sufficiently influential in auditory streaming (McDermott et al., 2011; Shinn-Cunningham et al., 2007; Lee and Shin-Cunningham, 2008), it remains unclear whether this influence is mainly due to “long-term knowledge” or “short-term knowledge”.

PREDICTABILITY

Another substantial influential factor is predictability (for a review see Bendixen et al., 2014). For instance, several studies found that temporal regularities stabilized and prolonged the duration of the perception of separate streams (Bendixen et al., 2010 and Devergie et al., 2010) and numerous following works provided support for a role of predictability in auditory streaming (Bendixen et al., 2013, 2014; Sohoglu and Chait, 2016; King and Walker, 2020). However, a better comprehension of the interactions between predictability, prior knowledge, familiarity and bottom-up cues is necessary to understand at which level of auditory streaming and how influential predictability is (Bendixen et al., 2014).

ATTENTION

Attention is thought to be a critical factor in auditory streaming. However, it is an ongoing debate to know whether it can modulate sound features representations and/or auditory streams. In other words, does attention operate only once the auditory stream has been

formed, which would mean that auditory stream formation could be a fully automatic process requiring no attention, or does it come into play at the level of stream formation, which would suggest that attention could be necessary for the formation of auditory streams?

Before going into this interesting debate, it is essential to deliver some details about what auditory attention is, *per se*. Although it is generally assumed that attention is key to perceive the world surrounding us, it remains a non-unified notion that covers a lot of different meanings (Oberauer, 2019). Giving a consensual definition of what is attention would be, thus, unrealistic. Some authors have even argued that this fragmented term should be abandoned for the benefit of the scientific community (e.g., Anderson, 2011, 2021; Hommel et al., 2019; Allport, 1993).

Attention can be seen as a function that deals with the limited amount of resources available to process information (Pashler, 1998). Importantly, this notion of a limited amount of resources, introduced by early theories of attention (Broadbent, 1958; Deutsch and Deutsch, 1963; Treisman, 1964), is now integrated in almost all the conceptions of attention and working memory.

Another conception is to think about attention as a general selection mechanism used to prioritize actions⁶ (Neisser, 1976; Wu, 2014, 2019; Watzl, 2017). This view was substantiated by early (Neumann and Allport, 1987) as well as more recent theories (Shinn-Cunningham, 2008; Chait et al., 2010; Shamma et al., 2011; Maddox et al., 2012; Mesgarani and Chang, 2012; Middlebrooks and Bremen, 2013) claiming that a competition between objects in the auditory scene occurs because only one object can be in the focus of attention at the time.

More recently, in the framework of predictive coding (Friston et al., 2006) it has been proposed that attention could be seen as a weighting mechanism related to the

⁶ The term *actions* refers to both bodily behavior and mental activities, such as puzzle solving.

optimization of the precision of one's expectation about the world (Hohwy, 2012; Feldman and Friston, 2010).

Finally, another interesting point of view is to consider attention as a system separated in three interconnected subsystems, each representing different attentional processes (Petersen and Posner, 1990, 2012; Mackie et al., 2013, Posner and Fan, 2008). The first system is in charge of maintaining optimal vigilance. The second, is responsible for selecting the most relevant information from various inputs within and across modalities (Petersen and Posner, 2012). The third detects and resolves conflict among competing mental processes and is thought to be top-down modulated and related to executive functions (Petersen and Posner, 2012) This systemic view of attention reconciles theories of selections and prioritization of information. The attentional system is viewed as hierarchical, with executive control network at the highest level and alerting and orienting networks at a lower level (Wang and Fan, 2007; Mackie et al., 2013). In addition, the control network is thought to be supramodal whereas the other networks may be modality specific (De Santis et al., 2007; Thiel and Fink, 2007; Arnott et al., 2004; Kong et al., 2012; Spagna et al., 2015).

Apart from theories and conception, attention can also be divided into different types. In reality, this distinction is questionable, as attention's type interact with each other and are potentially supported by the same neural networks (see for example Hahn et al., 2008 and Hommel et al., 2019 but see Pinto et al., 2013 for opposite results) but the most common division generally separates divided, selective and sustained attention.

Divided attention characterizes the allocation of attentional resources between several stimuli by rapidly switching the attentional focus (Parasuraman, 1998). Switches can be either intentional or unintentional. In the context of ASA, dividing attention between concurrent stimuli always comes with a cost and impairs performance compared to situations where attentional resources are fully devoted to the processing of one stimulus. This has been demonstrated by studies comparing conditions where participants are required to report the information presented in a target stream to conditions where they were asked to report two information encompassed in two streams presented simultaneously (e.g., Best et al., 2010; Gallun et al., 2007). Other studies have illustrated this *attentional switching cost* in situations where participants

had to switch between multiple speech streams (Best et al., 2008; McCloy et al., 2017; Lawo and Koch, 2014; Larson and Lee, 2013; Koch et al., 2011). In addition, unintentional switches of attention (i.e., attentional capture) may happen when a salient stimulus appears in the auditory scene (see Dalton and Hughes, 2014 for a review). A textbook example is when attention is attracted by one's own name (Moray, 1959). But this saliency may also be dictated by other factors such as violation of expectancy (e.g., Hughes et al., 2007; Nössl et al., 2012; Parmentier et al., 2011), abrupt changes in sound level of words (Holender, 1986), biologically relevant words (Parmentier et al., 2018) or inharmonicity (Bonin and Smilek, 2017).

However, it remains true that divided attention can be important in natural auditory scenes. For instance, to understand a conversation, attention must constantly be switched between the different speakers.

Selective attention can be seen as the opposite of divided attention because it refers to the ability to allocate attentional resources on a specific stimulus while ignoring irrelevant ones in the environment (Murphy et al., 2017). This type of attention is mainly top-down driven and dependent on inhibitory control (van Moorselaar and Slagter, 2020). Selective attention is particularly important in complex auditory scenes where effective processing of a target sound source requires segregating this target from the other irrelevant stimuli.

Auditory selective attention in the context of ASA was first studied by Cherry (1953) and Broadbent (1952, 1962) using the dichotic listening paradigm. In their experiment, participants were presented with different spoken words in each ear (see Driver, 2011 for a review) and instructed to selectively focus attention on the words presented in the designated ear while ignoring the words presented in concurrent one. They found that words in the concurrent ear were poorly recalled compared to the words in the selected ear. Their findings gave birth to the early theories of attentional filters (Broadbent, 1958, Deutsch and Deutsch, 1963; Treisman, 1963; Treisman et al., 1980). Later, Alho and colleagues (1987) exposed that the amplitude of the ERP attention effect increased when the frequency separation between attended and unattended streams increased (see also Alho et al., 1986a, 1986b; Hansen and Hillyard, 1980; Alain and Woods, 1994, Greenberg and Larkin, 1968). More recently, a huge body of work has demonstrated the impact of

selective attention on ASA. For instance, Carlyon and collaborators (2001) as well as Cusack and colleagues (2004) found that the construction of auditory streams (the “buildup”) was sensitive to selective attention. Snyder and colleagues (2006) found consistent results and showed reduced buildup-related neural activity for unattended stimuli compared to the attended one. And these results were corroborated by a multitude of other studies showing an effect of selective attention on auditory streams (e.g., Gutschalk et al., 2008; Thompson et al., 2011; Bidet-Caulet et al., 2007, 2010; Lipp et al., 2010; Billig and Carlyon, 2016; Chait et al., 2010; Lu et al., 2017; Shamma et al., 2011; Elhilali et al., 2009). Finally, in the last decade, several studies highlighted the fact that when attention is focus on a specific speech stream, its cortical representation in enhanced compared to when the attention is not directed to it (e.g., O’Sullivan et al., 2015; Zion Golumbic et al. 2013; Pasley et al., 2012; Puvvada and Simon, 2017; Mesgarani and Chang, 2012).

Although there is a consensus to say that top-down selective attention enhances stream segregation, it is still an open question to know whether selective attention is necessary to segregate streams (Shinn-Cunningham and Best, 2015). Interestingly, mixed evidence can be found in the literature. For instance, some works established that attention is necessary for stream segregation (e.g., Jones, 1976 ; Alain and Woods, 1997; Carlyon et al., 2001; Gutschalk et al., 2008; Lu et al., 2017; Molloy et al., 2019), others claim that scene analysis can occur without direct focus of attention (e.g., Macken et al., 2003; Sussman et al., 2007; Teki et al., 2011, 2016) and still others proposed a dual process, consisting of a pre-attentive segregation mechanism and an attention-dependent mechanism (Cusack et al., 2004; Snyder et al., 2006; O’Sullivan et al., 2015a). A possible global mechanism of stream segregation could be that when bottom-up cues are sufficiently informative, stream segregation occurs without attention but when this automatic mechanism is not sufficient, top-down attention is coming into play to help stream segregation (Shinn-Cunningham and Best, 2015; Sussman, 2017).

Finally, Sustained attention could be defined as the ability to allocate attentional resources to a task at hand during a certain amount of time (Fortenbaugh et al., 2017). Interestingly, sustained attention is less studied than divided or selective attention (Esterman and Rothlein, 2019) even if the ability to stay focused on a task is important in

many aspects of our life (e.g., Edkins and Pollock, 1997; Taylor-Phillips and Stinton, 2019).

Sustained attention is highly related to inhibitory control because keeping attention focused on a stimulus requires inhibition of external as well as internal distractors (Stuss, et al. 1995; Clayton et al., 2015). Therefore, it may appear obvious to say that sustained attention is important in ASA because it is related to inhibition of distracting sound sources but several works have found evidence for its importance in ASA. As proposed by Best and colleagues (2008), exercising sustained attention on an auditory object may reinforce attentional selectivity. This proposition is in line with studies showing that segregation can occur automatically but that the buildup of streams refines through time and therefore depends on sustained attention (e.g., Cusack et al., 2004; Snyder et al., 2006; Snyder and Alain, 2007a; Winkler et al., 2005; Sussman, 2017). In addition, some studies have found involvement of sustained attention in speech comprehension (Tierney et al., 2019; Le Gal de Kerangal, 2020) as well as in dichotic listening (Helge Johnsen et al., 2002; Hommet et al., 2010; Jäncke et al., 2003). Other interesting evidence is coming from studies examining age-related effects of decreased scene analysis capacities. For instance, Panek and Rush (1981) found that the ability to sustain attention in a dichotic listening situation was reduced in older adults. Snyder and Alain (2007) suggested that older adults impairment of speech comprehension in complex auditory scenes could be partly due to a deficit in concurrent sound segregation induced by a deficit in sustained attention.

Overall, these elements support the implication of sustained attention in the ability to analyze complex auditory scenes even if sustained attention has been far less studied than selective attention, in the context of complex auditory scenes.

In summary, attention(s) can be conceptualized in many ways and it seems to be an important factor impacting ASA. Even if we do not completely understand its role on stream formation and segregation, recent theories as well as advances in the comprehension of sound's cortical representations and brain imaging techniques started to bring some elements to answer the question.

That being said, beside the ones mentioned previously, other top-down factors can also impact ASA. This is for example the case of motivation or fatigue that could impact

attention and therefore auditory streaming. Another key set of cognitive processes, closely related to attention, can also exert a critical influence on ASA, the *executive functions*.

EXECUTIVE FUNCTIONS

Executive functions are important in many aspects of our daily life because they are essential components of higher-order processes such as problem solving, planning or reasoning (Collins and Koechlin, 2012; Lunt et al. 2012; Diamond, 2013, 2020) and their involvement has been shown for instance in job success (Bailey, 2007) and school achievement (e.g., Borella et al., 2010; Duncan et al., 2007; Gathercole et al., 2004; Gilmore and Gragg, 2014; Zelazo et al., 2016). One can conceptualize executive functions as a set of top-down mental processes necessary when concentration and attention are required and when it is impossible or insufficient to rely on automatic schemas such as, for instance, in an unpredictable situation (Burgess and Simons, 2005; Espy, 2004; Miller and Cohen, 2001; Diamond, 2013, 2020). Therefore, their use is effortful as it requires attentional resources. There is a consensus to say that executive functions can be divided into three core functions: Cognitive Flexibility, Working Memory (WM) and Inhibitory Control (IC) (Miyake et al., 2000; Diamond, 2013, 2020; Lehto et al., 2003).

Cognitive flexibility is probably not the most important executive function for ASA. Although one can think about its role in conversations which require cognitive flexibility to smoothly manage speaking turns. In addition, cognitive flexibility is probably based upon both WM and inhibitory control (Diamond, 2013). Therefore, we will not discuss this function here but the interested reader can find useful information in the reviews of Diamond (2013, 2020).

WORKING MEMORY

WM is way more important for ASA and auditory streaming in particular. As a consequence, before providing elements about the reasons why it is so critical for auditory streaming, we will go into more details about how it is conceptualized and what are the classical paradigms to measure it.

Similar to the notion of attention, the notion of WM is not completely unified. In a recent review, Cowan (2017) listed no less than 9 definitions depending on the theoretical framework in which working memory was considered. However, as mentioned by Adams

and colleagues (2018) a generic definition, independent of any theoretical framework, could be that WM is “*a system of components that holds a limited amount of information temporarily in a heightened state of availability for use in ongoing processing*”. In light of this definition, it appears that WM is probably crucial for making sense of anything that is dynamic in time such as reading (see for example Savage et al., 2007 for a review of WM and reading difficulties) or having a conversation (e.g., Bender, 2004; Waters and Caplan, 1996). WM is also involved as long as one needs to manipulate information (e.g., making connections of thoughts and concepts or doing mental calculation) and is thought to be involved in numerous higher-order cognitive functions such as reasoning and problem solving (e.g., Engle, 2002).

Since this early *modal model* of Atkinson and Shiffrin (1968) many models of WM have flourished. One of the most influential ones was the *multi-component model* developed by Baddeley and Hitch (1974, then refined in 1986 and 2000). In its former proposition, the model was composed of three modules, one master module named the *central executive* and two slave modules, *the phonological loop* and *the visuospatial sketchpad*. The slave modules are attention-free, active storage units and they are responsible of maintaining the information, arriving from the environment, in the system. They are modality specific; *the phonological loop* is only in charge of the verbal information while *the visuospatial sketchpad* is in charge of the visual and spatial information. The information stored in these modules need active rehearsal to be maintained because they are subject to decay. One of the proposed mechanisms is the *subvocal repetition* which is an internal repetition mechanism allowing the reactivation of the verbal information in *the phonological loop* module. In addition, stored information is conditional to interference from the processing of newer information of the same modality that enters the system. Yet, as the slave modules processes information independently, the processing of visuospatial information should not interfere with the processing of verbal one. This has been demonstrated by studies who found that the recall of visual pattern was not affected by a concurrent task in the verbal modality (e.g., Cocchini et al., 2002; Bayliss et al., 2003; Logie et al., 1990) and vice versa. *The central executive* is not a storage module but it can be thought of as a supervisory module. For Baddeley (1996), it has multiple roles, including the control and the allocation of the limited attentional resources to the different slave modules but also the initiation of rehearsal process. A fourth module was

added more recently (Baddeley, 2000) named *the episodic buffer* which is thought to be a storage system involved in the maintaining of the combination of visuo-spatial and verbal information but also of abstract information.

Another significant model was the embedded-process model of Cowan (1988, then refined in 1999 and 2005) developed in response to Baddeley's model. In reality, the two models are not fundamentally different as mentioned by Cowan (2005). They are both multi-component models, as the embedded-process model includes a central executive, a sensory buffer and a long-term memory (LTM) store. However, the embedded-process model is more generic because it proposes a common system for phonological, visuospatial as well as for information of other modalities, such as tactile ones, and it includes both short-term memory (STM) and LTM. In Cowan's model, the information coming from the environment first passes through the *sensory buffer* where they activate *memory traces*, which are stored in LTM. These memory traces could either be *inactivated* (they remain in LTM and are not consciously accessible), *temporarily activated* (they pass from LTM to STM and are available to consciousness) or *activated* in STM (they are in STM and in the attentional focus). The fate of the memory traces depends upon the attentional focus because when a memory trace is temporarily activated in STM, it is subject to time decay. The only way to avoid the memory trace to disappear from STM, and consequently to become inaccessible to consciousness, is to activate it by "moving it" inside the attentional focus. This attentional focus is under the control of the *central executive* module that will allocate the attentional resources depending on bottom-up (e.g., salience of a stimulus) and top-down factors (e.g., goals, instructions). Additionally, Cowan's model postulated that the attentional focus is limited to four items (e.g., Chen and Cowan, 2005, 2009; Cowan et al., 2004). Interestingly, the information in working memory is subject to interference from novel information entering the system independently of its modality. It thus makes the assumption of a domain-general WM system. This claim is supported by several experiments founding cross-modal interferences (e.g., Sauls and Cowan, 2007; Morey and Cowan, 2005; Sirevaag et al., 1989). Overall, in the Embedded-processes model, the WM is more a process embedded in LTM and underlied by attention than a system *per se*.

Slightly differently, Ericsson and Kintsch (1995) suggest that the WM could be a system of the LTM, called *Long-Term Working Memory* (LTWM), dedicated to the storage of

information about specific knowledge stored in LTM. This information stored in LTWM is thought to be retrieval cues that are required when one needs to retrieve specific knowledge stored in LTM. The unity of WM and LTM is also supported by other researchers claiming, among other, that time decay does not affect recall (e.g., Nairne, 2002; Jalbert et al., 2011; Neath and Brown, 2012).

Another “less influential” model of working memory, the Time-Based Resource Sharing (TBRS) model from Barrouillet and colleagues (2004 refined in 2007 and 2009) claims that the mechanisms of maintaining and processing information in WM operate sequentially and are dependent on the same amount of attentional resources, which is limited. Therefore, as maintaining and processing do not operate simultaneously, attention has to be allocated successively to one process or the other, by rapid switches of attention. These attentional switches are important because memory traces deteriorate when they are not in the attentional focus so they need to be constantly reactivated. Markedly, in the TBRS model, recall performances are thought to be dependent on the time during which attentional resources are devoted to the information processing and not necessarily to the processing time *per se* (e.g., Barrouillet et al., 2004; Barrouillet et al., 2007). Finally, as for the Embedded-processes model, WM is conceptualized as an amodal system because attentional refreshing operates on any kind of information (e.g., verbal, tactile, auditory, etc.).

Several other models of WM are available nowadays (e.g., Logie, 2016; Oberauer and Lin, 2017).

What seems consistent however, in (almost) all the models, is the essentiality of attention in the mechanisms driven by the WM (see for example Adams et al., 2018 and Oberauer, 2019 for discussions about the role of attention in WM models) no matter if these mechanisms are modality-specific or generic.

There are a considerable variety of tasks used by researchers to measure WM, ranging from very simple to very complex ones. One of the most widely used types of WM assessment is the so-called *complex span task (CST)* (Daneman and Carpenter, 1980; Diamond, 2013; Redick et al., 2012; Foster et al., 2014). In such a task, participants are presented with a sequence of to-be-remember items (e.g., letters, digits, words) and they have to perform a distracting task, like solving a math problem (Kane et al., 2004;

Unsworth et al., 2005) or doing mental rotations (Kane et al., 2004; Harrison et al., 2013), between the presentation of each item. After each *to-be-remember + distracting task* sequence, participants are asked to recall the to-be-remember items in serial order. The most common metric is the summation of the number of to-be-remember items correctly recalled (i.e., in the correct order) but one can also find metrics reflecting the number of errors in the distracting task (processing error score) for instance. Many parameters of these tasks can be modified in order to increase the difficulty such as the type of distracting task, the time allocated to process the distracting task or the number of elements to process within it (Barrouillet et al., 2004, 2007; Gavens and Barrouillet, 2004). However, as noticed by Diamond (2013), CSTs require more than just holding and manipulating information and are consequently measures of EF rather than pure measures of WM.

Another frequently used family of tasks is the N-back tasks (Owen et al. 2005, Verhaeghen and Basak, 2005) in which participants are presented with a sequence of items (e.g., words, letters, digits) and have to decide whether the currently presented item corresponds to the one presented N items before. Typical scores are the number of errors and the response times which are thought to increase with the size of N. In general, the N is manipulated across blocks to evaluate the effect of memory load on behavioral performances and neural correlates. It is also possible to include lure trials or to incorporate the N-back task into a *dual-task paradigm* where attention has to be divided between to N-back tasks presented simultaneously (e.g., Jaeggi et al., 2010). However, this type of WM measurement suffers from the same problem as the CSTs, they involve processes that go beyond the simple holding, manipulation and updating mechanisms associated with WM such as decision, selection (Jonides et al., 1997) and conflict monitoring (Oberauer, 2005, Kane et al., 2007).

Finally, the third family of WM tasks is the Simple Span Tasks (SSTs) which are massively used in neuropsychology (Hilbert et al., 2015). SSTs generally encompass Forward, Backward and Reordering span tasks. In these tasks participants are presented with a sequence of items (generally digits but it can also be words or letters) and they have to immediately repeat back in the same order in which they heard them (Forward Span Task), in the reverse one (Backward Span Task) or in a specific pre-established order (e.g. numerical; Reordering Span Task). The length of the sequence increases until the

participant makes two errors in a row. The longest correctly reported sequence reflects the WM span (the maximal amount of information one can manipulate and recall). Although forward span tasks are considered as measures of STM, because it does not require manipulation (e.g., Diamond, 2013; St.Clair-Thompson, 2010), there is an ongoing debate regarding whether backward and reordering span tasks are measures of STM or WM, despite the fact that the sequence transformation induced by these tasks requires attentional resources (Alloway et al., 2006; Elliot et al., 1997). Consistent with the idea that backward and forward digit span tasks tackle different constructs, several studies have found that backward digit span task and more complex WM measures such as CSTs load onto a common factor whereas it was not the case for forward digit span task (Alloway et al., 2004, 2006; Gathercole et al., 2004; Oberauer et al., 2000). Therefore, backward as well as reordering span tasks could be used to measure WM even if there is no agreement about it in the literature.

As briefly mentioned earlier, WM seems to be involved in anything that is evolving in time. Therefore, it could seem obvious that WM should be involved in auditory streaming, where effective following of an auditory stream requires making connections between dynamic auditory events. Consistent with this idea, a study on Alzheimer's disease patients found that performance on auditory streaming was influenced by non-verbal WM capacity measured using backward digit span task (Goll et al., 2012). The authors proposed that working memory impairment in patients could explain their poor performance in grouping because this process is "*dependent on the capacity to track and to bind auditory information evolving in time*" which is partly provided by WM. The importance of WM in auditory streaming received further neuroimaging support by the review of Christison-Lagay and colleagues (2015) which explain that the ventrolateral prefrontal cortex (vlPFC) implicated in the processing of several features of auditory stimulus (e.g., content, identity or meaning) is also thought to reflect auditory WM (Plakke et al., 2012). These elements suggest that WM could be involved in auditory object representation.

The role of WM in auditory streaming has also been substantiated by multiple studies showing its implication in speech-in-noise perception, a situation thought to engage ASA processes. For instance, Akeroyd (2008) revealed that, across 20 studies, WM was the best predictor of speech-in-noise perception among several other cognitive measures.

Multiple other works found that greater WM capacity was related to enhanced speech-in-noise perception (Besser et al., 2013; Sorqvist and Ronnberg, 2012; Souza et al., 2015; Escobar et al., 2020; Parbery-Clark et al., 2009a; Ingvalson et al., 2015; Gordon-Salant and Cole, 2016) and dichotic listening (e.g., James et al., 2014; Conway et al., 2001; Colflesh and Conway, 2007).

Finally, WM was also related to polyphonic music listening (see Schulze and Koelsch, 2012 for a discussion about WM in music). Janata and colleagues (2002) for instance, underline that listening to polyphonic music recruit neural networks involved in WM, such as the parieto-frontal network. Findings supported by works that have found similar results (e.g., Gaab et al., 2003; Koelsch et al., 2009; Schulze et al., 2011a; Burunat et al., 2014).

In light of the above, it sounds that WM plays a role in auditory scene analysis. However, most of the abovementioned studies addressed the issue of auditory WM using verbal material. Therefore, WM's exact place in music listening is not completely clear.

INHIBITORY CONTROL

Similarly to WM, inhibitory control is a key executive function for ASA. We will first dive into its different conceptualizations and the classical paradigms used to measure it. Afterward, we will provide insights about the reasons why inhibitory control is important for auditory streaming.

Just the once will not hurt, *inhibitory control* (IC) can be classified as an umbrella term (Diamond, 2013; Tiego et al., 2018) but a generic definition could be found in Diamond's review about executive functions (2013): "*inhibitory control involves being able to control one's attention, behavior, thoughts, and/or emotions to override a strong internal predisposition or external lure, and instead do what's more appropriate or needed*". Consequently, IC is critical for day-to-day life because it allows us to control how we act, react and behave instead of being driven by habits or external stimuli.

One can distinguish between two main categories of IC: inhibitory control of responses (response inhibition) and inhibitory control of attention (attentional inhibition). Response inhibition refers to the ability to control one's behavior (e.g., inhibiting prepotent motor responses) while attentional inhibition concerns the ability to resist interference from both external (e.g., an irrelevant speaker) and internal distractors (e.g.,

unrelated thoughts) (Diamond, 2013, 2020; Tiego et al., 2018 but see Howard et al., 2014 for classification based on *automatic versus effortful inhibition*). However, it is interesting to note that some authors remain skeptical about this dichotomy (e.g., Friedman and Miyake, 2004; Kane et al., 2016; Gandolfi et al., 2014).

The motor side of response inhibition has been largely studied using the *stop-signal paradigm* (Schall et al., 2017; Verbruggen and Logan, 2009). In such a task, participants have to press a specific key depending on the presented stimulus (e.g., “A” if it is a square and “P” when it is a circle) and withhold their response when the stimulus is followed by a *stop-signal* (e.g., a tone). Several studies found that the ability to inhibit the response was dependent on the delay between the presented stimulus and the stop-signal (e.g., Vince, 1948; Lappin and Eriksen, 1966) because this *stop-signal delay* represents the time available to detect the stop signal and revoke the response before its execution. And this effect has been observed for different types of material, modalities and population (Schall et al., 2017).

These observations gave birth to the dominating model of response inhibition called the Independent Race Model which proposed to see the inhibition of motor responses as a race between a STOP process and a GO process (Ollman, 1973; Logan, 1981). Therefore, it is thought that when the STOP process finishes before the GO process, response inhibition is accomplished. Conversely, when the GO process finishes before the STOP process, response inhibition is unaccomplished. Since the first formal implementation of this model by Logan and Cowan (1984) multiple neurophysiological works have found results consistent with the hypothesis of a competition between stop and go processes (e.g., Hanes et al., 1998; Pouget et al., 2017; Paré and Hanes, 2003; Mallet et al., 2016). As a consequence, efforts have been made to improve the Independent Race Model by incorporating neural circuitry (Frank, 2006; Wiecki and Frank, 2013) and richer models have emerged such as the Interactive Race Model (Boucher et al., 2007; Wong-Lin et al., 2010; Ramakrishnan et al., 2012) or the Spiking Network Interactive Race Model (Lo et al., 2009).

Other tasks have been classically used to measure response inhibition such as the Go/NoGo tasks (Drewe, 1975; Garavan et al., 1999; Picton et al., 2007) or the anti-saccade task (Hallet, 1978; Munoz and Everling, 2004). In the Go/NoGo tasks, in general, there are

75% of go stimuli and 25% of no-go stimuli. This peculiar configuration induces a quasi-automatic motor response and consequently, the engagement of inhibitory processes to suppress the motor response is strong (Simmond et al., 2008). In the antisaccade tasks, participants are visually displayed with a central cross that is replaced by a sudden target and have to look at the contralateral position of the target. This implies to inhibit the automatic response to look toward the target (i.e., a pro-saccade), that acts as a salient stimulus.

On the attentional side of IC, one can make the distinction between resisting interference from external or internal distractors. Inhibition of internal distractors can be viewed as the ability to resist (inhibit) internal/unwanted thoughts or mind wandering. Mind wandering is defined as “*a shift in the contents of thought away from an ongoing task and/or from events in the external environment to self-generated thoughts and feelings*” (Smallwood and Schooler, 2015). While mind wandering is thought to be beneficial for the system to generate new and creative thoughts (Baird et al., 2012; Ruby et al., 2013b) and cognitive rest (Baird et al., 2010; Ruby et al., 2013a) it remains disruptive for a certain number of tasks because it decreased awareness of the external environment (Smallwood et al., 2003; Smallwood and Schooler, 2006). For example, it has been found that mind wandering interferes with comprehension during reading (Dixon and Bortolussi, 2013; Feng et al. 2013; Franklin et al. 2011; Jackson and Balota, 2012; McVay and Kane, 2011; Schooler et al. 2004; Smallwood et al., 2008, 2013; Unsworth and McMillan, 2013), performance during tasks requiring high level of executive control (Mrazek et al., 2012) but also, and it is more problematic, with reaction times, environment monitoring and velocity during simulated driving situations (Yanko and Spalek, 2014; He et al., 2011).

To investigate mind wandering and unrelated thoughts, researchers are using *Experience Sampling* (ES) methods (Kahneman et al., 2004). The most common ES technique is the *probe-caught method* (Smallwood and Schooler, 2006) in which participants are randomly or quasi-randomly interrupted during the task and probed about the content of their thoughts (e.g., related or unrelated to the task at hand). In the same vein, in the *self-caught method* (Smallwood and Schooler, 2006) participants are asked to spontaneously report when they notice their mind wandering. The main problem with these two methods is that they interrupt the natural dynamic of the task at hand (Barron

et al. 2011; Smallwood et al. 2012) but also the natural dynamic of attentional processes because probes periods can act as rest periods, allowing attentional resources restoration. Finally, in the *retrospective method* data is collected at the end of the task. This method is less disruptive as it preserves the attentional dynamic but it requires participants to remember the number of times they catch their mind wandering during each trial which can suffer from forgetting and mental-aggregation biases (Kane et al., 2021).

Because it is hard to cause the mind to wander and because it is an internal phenomenon with very few external manifestations, it remains difficult to measure without using subjective report techniques. However, the disengagement of attention from events taking place in the here and now, occurring during mind wandering, named *perceptual decoupling of attention*, has been investigated using objective methods such as EEG and pupillometry (e.g., Smallwood, 2011). For instance, several studies found that the amplitude of the P300, an EEG marker of task-related attention (Polich, 1986), as well as other components indicating sensory processing of auditory and visual information, were reduced during periods of mind wandering (Barron et., 2011; Kam et al., 2011, 2014; Macdonald et al., 2011). The use of objective measures seems therefore to be both achievable and useful to understand what is mind wandering and how it interacts with world perception.

The other facet of attentional inhibition is defined as the ability to resist interference from stimuli in the external environment. This ability enables human beings to focus attention on a specific target without devoting attention to other stimuli. A textbook case of external distractors inhibition could be the cocktail party scenario. When someone wants to focus attention on a target auditory source, voluntary inhibition of the concurrent sources is required. It is therefore driven by one's goals or intentions⁷. In the auditory

⁷ Notice that the thin frontier separating inhibition of external distractors and attention often induced some confusions between the terms, some studies even using interchangeably *attentional inhibition*, *selective attention*, *focused attention*, *top-down attention* or *attention control*. In a comprehension purpose, the term *attentional inhibition* will refer to *the ability to resist interference from external or internal distractors*

modality, Bidet-Caulet and colleagues (2007) demonstrated that selective attention was underlied by two different mechanisms depending on the conditions. When the auditory scene was encompassing only one sound, attention operated via a facilitation mechanism by enhancing the neural representation of this sound. However, when the auditory scene was encompassing two different sounds, attention operated via an inhibitory mechanism by reducing the neural representation of the irrelevant sound. This last result was consistent with previous studies highlighting this inhibitory mechanism (e.g., Ghatan et al., 1998; Alho et al., 1994; Donald, 1987; Michie et al., 1990, 1993; Alain and Woods, 1994) and was replicated later (Bidet-Caulet et al., 2010; Chait et al., 2010).

Nevertheless, an interesting question in the context of complex auditory scenes is: “does the formation of auditory streams that supposedly occurs in the absence of attention based on bottom-up cues repose on pre-attentive inhibition mechanisms?” (Sussman et al., 2007; Macken et al., 2003; Teki et al., 2011, 2016 but see Carlyon et al., 2001; Gutschalk et al., 2008; Lu et al., 2017 and Molloy et al., 2019 for opposite results)

A possible element of the answer may be found in the visual modality literature where inhibition of salient external distractors has been widely studied (Gaspelin and Luck, 2018; van Moorselaar and Slagter, 2020). Recent advances in this domain demonstrated that this inhibitory mechanism does not necessarily require attention (Gaspelin and Luck, 2018), even if voluntary inhibition can enhance the processing of relevant information (Battistoni et al., 2017; Posner, 1980). Moreover, some evidence suggests that prior knowledge about statistical regularity of distractors could explain this preattentive inhibitory mechanism (Chelazzi et al., 2018; Noonan et al., 2018; Gaspelin and Luck, 2018). Therefore, it seems that, in the visual modality, two mechanisms of inhibition are at work, one operating without attention and another one, driven by one’s goals, that support selective attention. Is this also the case in the auditory modality? Even if the parallel between the two situations is not straightforward one can think it may be possible that inhibition in the auditory modality could be controlled by bottom-up and top-down mechanisms depending on the conditions.

whereas the term *selective attention* will refer to the *ability to concentrate attention on a specific stimulus*.

Besides the specific cases previously discussed, the general construct of attentional inhibition as an EF is commonly measured using *Stimulus-Response Compatibility (SRC) tasks*. These are forced-choice reaction time tasks where the participants have to respond to a stimulus while ignoring (inhibiting) irrelevant information. The most used SRC tasks (Diamond, 2013) are the *Simon Task* (Simon, 1969; Simon and Berbaum, 1990; Hommel, 2011), the *Flanker Task* (Eriksen and Eriksen, 1974; Mullane et al., 2009), and the *Stroop Task* (Stroop, 1935; MacLeod, 1991).

In the Simon task, participants are visually presented with stimuli appearing either on the left side or on the right side of the screen and are instructed to press on the right or on the left depending on a non-spatial dimension of the stimuli (e.g., blue = “left”, green = “right”). It was observed that although the location of the stimulus is not relevant, participants are faster and make less error when the non-spatial dimension of the stimulus is congruent with its location (e.g., when a blue stimulus appears on the left side of the screen) compared to the opposite case. This observation was named the *Simon effect* and reflects the inhibition of the location of the stimulus (i.e., the irrelevant information) necessary in the incongruent trials. Scores in such a task are the difference of reaction time between congruent and incongruent trials as well as the accuracy of congruent and incongruent trials. Interestingly, the Simon effect has been documented in the visual modality (Simon, 1969; Hommel, 1993), the auditory modality (Craft and Simon, 1970) and the somatosensory modality (Hasbroucq and Guiard, 1992).

In the Flanker task participants are required to report the identity of a centrally presented stimulus (e.g., pressing “left” or “right” if an arrow is pointing to the left or to the right, respectively) and ignore (inhibit) flanking distractor stimuli surrounding it. The flanking stimuli can be either congruent or incongruent with the target stimulus. The main observations derived from this task are slower response times and a lower accuracy when the flanking stimuli are incongruent (Botvinick et al., 2001; Eriksen and Eriksen, 1974). This was named the *Flanker effect* and expresses the involvement of top-down control necessary in the incongruent trials. Scores computed from the Flanker task are similar to those of the Simon task. Several studies found an impact of expectations on *flanker effect* (Ghinescu et al., 2016; Wühr et al., 2018; Russeler et al., 2003; Bulger et al., 2020) suggesting the existence of an expectation-mediated inhibition mechanism potentially consistent with the hypothesis of a preattentive inhibition mechanism in the

visual modality (Noonan et al., 2018). That being said, interference effect occurring during the incongruent trials in the flanker task is thought to be mainly due to *stimulus-response conflict* (Eriksen and Eriksen, 1974; Eriksen and St James, 1986; Eriksen, 1995; van Veen et al., 2001) and not to *stimulus-stimulus conflict* which means that the performance could reflect response inhibition instead of attentional inhibition (Tiego et al., 2018).

Finally, the most clinically used measure of attentional control is the Stroop task. In the visual color-word Stroop task (i.e., the most common one, see Stroop, 1935), participants are presented with three boards they have to read as fast and accurately as possible. The first one is filled with color words printed in black ink and participants have to read the name of the written colors. The second one is filled with color patches and participants are required to name the colors. These two boards are the congruent conditions. The last board is filled with color words printed in inconsistent colors (e.g., the word “blue” printed in red) and the participant needs to name the color of the inks and not to read the words. This last board is the incongruent condition. Classically, reading time is lower and accuracy is higher for congruent than for incongruent condition because the incongruent condition needs the participant to inhibit the automatic reading of the words to name the color. This difficulty in the inhibition of this automatic process is named the *Stroop effect* (Stroop, 1935). In general, an interference score is computed, reflecting attentional inhibition ability *per se*, as the reading time difference between the congruent and incongruent conditions. At the margin of the traditional visual version, auditory versions of the Stroop task have also been developed (e.g., Morgan and Brandt, 1989; Jerger et al., 1988; Most et al., 2007; Donohue et al., 2012; Danneels et al., 2020; Knight and Heinrich, 2017; Strouwen et al., 2016; Wurm et al., 2004; Zekveld et al., 2020). Here, participant is required to respond as fast and accurately as possible to a particular perceptual feature of a word (e.g., voice pitch, speaker gender or stimulus location) while ignoring (inhibiting) other irrelevant perceptual information (generally the word in itself such as “low” pronounced at a high pitch or “left” heard in the right ear). Performance in auditory Stroop appears to be analogous to the visual version (e.g., Jerger et al., 1988; Morgan and Brandt, 1989; Knight and Heinrich, 2017; Zekveld et al., 2020) and scores remain the same as for the visual paradigm. Strikingly, a Stroop effect has also been documented for automatic note processing in musicians (Grégoire et al., 2013). Notice that, contrary to

Flanker task, interference effect in Stroop task is mainly due to *stimulus-stimulus conflict* which means that such a paradigm probably measures attentional inhibition (Simon and Berbaum, 1990; Zhang and Kornblum, 1999; Milham et al., 2001; Egner et al., 2007; Tiego et al., 2018).

Interestingly, several studies have compared visual and auditory Stroop in order to know if attentional inhibition was modality dependent. Some studies found similar neural activation patterns, correlated behavioral responses (Roberts and Hall, 2008) and similar neurophysiological responses for the two modalities (Donohue et al., 2012). Another one found that visual and auditory Stroop measures were uncorrelated (Knight and Heinrich, 2017). These contradictory results are confusing and urges for further exploration.

Finally, one thorny issue of all the above-mentioned tasks concerns their scoring method. In reality, like measures of sustained attention, and no matter the modality, the basic score usually computed (i.e., RT congruent - RT incongruent) does not take into account the Speed-Accuracy-Trade-Off (SATO) inherent in tasks requiring participants to be fast and accurate at the same time (Vandierendonck, 2021). Yet, it is possible to control SATO by incorporating both RT and error rates into a single metric (e.g., Gartner and Strobel, 2020, see also Vandierendonck, 2021 for a review on SATO).

Attentional inhibition seems to be essential to selectively listen to a source in a multi-source environment because inhibition of concurrent streams is necessary. This claim is underlined by previously evoked studies having suggested that an attentionally-driven inhibition mechanism, reducing the neural representation of the irrelevant sound, was at work when one needs to selectively listen to an auditory source in presence of a concurrent one (Ghatan et al., 1998; Alho et al., 1994; Donald, 1987; Michie et al., 1990, 1993; Alain and Woods, 1994; Bidet-Caulet et al., 2007, 2010; Chait et al., 2010). Support for this inhibition mechanism in auditory scene analysis can also be found in studies having explored neural oscillations, who also suggest that an inhibitory mechanism reduces the neural representation of distracting information (Fuxe and Snyder, 2011; Strauß et al., 2014).

Behavioral studies also show involvement of attentional inhibition in speech-in-noise listening (Knight and Heinrich, 2017; Campbell et al., 2020). Noticeably, a meta-analysis from Dryden and colleagues (2017) highlighted that the correlation between attentional

inhibition and speech in noise comprehension was of .34 across 6 studies. More surprisingly, behavioral measures of attentional inhibition in the visual modality have also been related to speech in noise comprehension (Knight and Heinrich, 2017; Janse, 2012; Sommers and Danielson, 1999). Finally, attentional inhibition is also thought to be involved in dichotic tasks (e.g., Hugdhal et al., 2009; Falkenberg et al., 2011).

Another evidence of the importance of attentional inhibition can be found in cocktail-party-like scenarios. In such situations, concurrent auditory streams may become salient for many reasons (e.g., a sudden volume augmentation) and capture attention (Dalton and Hughes, 2014) resulting in the loss of the source of interest. Switching back to the source of interest can take some time. Time during which crucial information could potentially be lost (e.g., missing the principal argument during an important conversation).

ATTENTION, WORKING MEMORY AND INHIBITORY CONTROL: A SPECIAL CASE?

Through the previous sections, we pointed to the fact that the notions of WM, IC and attention are sometimes confused in the literature (e.g., attentional inhibition and selective attention). This confusion seems to arise because these three notions are completely intertwined, sometimes leading to difficulties to properly disentangle them.

In reality, WM and IC need one another and co-occur (Diamond, 2013). For instance, knowing what to inhibit requires holding a goal in working memory. And to properly hold, manipulate and update information in WM, one needs to resist distractions. IC is sometimes viewed as a protective mechanism that protects the mental workspace from overload by suppressing no-longer-relevant information in WM (Hasher and Zacks, 1988; Hasher et al., 2007). And several conceptions of WM, considered IC as one of its sub-component or mechanisms. In Baddeley's model (1996), for instance, one of the many functions of the central executive include inhibitory control and in Cowan's model (2001), the maintenance of information in WM is ensured by the ability to block out (inhibit) distracting information.

This close link between WM and IC receives credit from empirical evidence showing that increasing workload in WM increases interference from salient task-irrelevant stimuli (e.g., Unsworth et al., 2004; Lavie and De Fockert, 2005; Burnham et al., 2014). Several studies have also highlighted a relation between WM capacity and interference effect in

the Stroop task (Long and Prat, 2002; Kane and Engle, 2003; Kiefer et al., 2005; Meier and Kane, 2013). Moreover, it has been suggested that both IC and WM rely upon common neural networks (De Fockert et al., 2001; Kane and Engle, 2003). For instance, McNab and colleagues (2008) found common activations in the right inferior frontal gyrus, the right middle frontal gyrus and right parietal regions during several WM and inhibitory control tasks.

Attention is also fundamentally linked to both WM and IC as evidenced by shared neural activations in prefrontal-parietal regions (e.g., Awh et al., 2000; Awh and Jonides, 2001; Gazzaley and Nobre, 2012; Ikkai and Curtis, 2011; LaBar et al., 1999; Nobre and Stokes, 2011). In addition, almost all the models of WM are talking about “a limited amount of attentional resources”. Of course, distinction about the type of processes requiring attentional resources (e.g., processing, storage, control of attention) can be made, depending on conceptions of what WM and attention are (Oberauer, 2019). However, this means that the limited WM capacity could be partly explained by the limited attentional resources.

It appears obvious to say that attention and IC are linked because selective attention is based on the ability to inhibit irrelevant stimuli. But this link is supported by direct evidence such as the above-mentioned studies who found the existence of an attentional-inhibitory mechanism reducing the neural representation of irrelevant sounds (e.g., Bidet-Caulet et al., 2007, 2010; Chait et al., 2010; Lu et al., 2017) or studies in visual modality showing the role of inhibitory mechanisms in attentional capture (e.g., Gaspelin and Luck, 2018) and selective attention (van Moorselaar and Slagter, 2020).

Another illustration of this link can be found in the sustained attention literature. Indeed, the most common measure of sustained attention requires inhibition of rare target stimuli (Esterman and Rothlein, 2019). This indicates that performance in a sustained attention task is partly driven by the ability to inhibit the induced prepotent response. This was confirmed by a latent variable analysis which found that response inhibition and slower response time in a sustained attention task, reflecting lapses of attention, was related (Unsworth et al., 2010). In addition, several works have shown correlation between performance to sustained attention tasks and mind wandering (Unsworth and

Robison, 2013; Kucyi et al., 2016; smallwood et al., 2004) which could be seen as a failure of inhibiting unrelated thoughts.

Finally, under the framework of *cognitive control* (Posner and Snyder, 1975), attention and IC are intimately joined because attentional functions are thought to be involved in the resolution of conflict by inhibiting irrelevant information (Mackie et al., 2013). In a complex auditory scene, the orienting system will direct attention to the location of the source of interest, then the executive control of attention system will inhibit the concurrent sources (Spagna et al., 2015) allowing for an efficient attentional focus on the source of interest.

SUMMARY

In summary, one can say that the ability to analyze a complex auditory scene is a fundamental day-to-day skill of human beings because it allows us to communicate with our peers. This ability is a perfect blend between automatic and attentionally-driven processes which interact with executive functions and more particularly with WM and IC.

MUSICAL EXPERTISE

OVERVIEW

In this section, we will explain why musical expertise is interesting in the context of ASA. First, elements about the potential impact of musical expertise on executive functions and attention are provided. The still controversial aspect of the enhancement of higher order cognitive function with musical training is also discussed. Then, some direct and indirect evidence for the influence of musical expertise on ASA are presented.

WHY IS MUSICAL EXPERTISE INTERESTING?

Musical expertise is thought to be an important factor modulating cognitive functioning, mainly because learning to play an instrument requires multiple sensory, motor and cognitive operations, leading to both structural and functional changes in the brain (Jäncke, 2009; Hannon and Trainor, 2007). An important amount of work has shown evidence for brain plasticity in musicians (see Herholz and Zatorre, 2012 for a review). Naturally, some studies point out changes in the auditory system including the brainstem (e.g., Musacchia et al. 2007; Wong et al., 2007) or the primary and secondary auditory

cortices (e.g., Bermudez et al., 2009; Schneider et al., 2002; Gaser and Schlaug, 2003; Foster and Zatorre, 2010) but several studies also found changes due to musical training in the sensorimotor network (e.g., Bengtsson et al., 2005; Schlaug et al., 1995; Hutchinson et al., 2003; Bangert and Schlaug, 2006; Hyde et al., 2009; Luo et al., 2012) or even in brain associative regions and regions engaged in syntactic processing (Herholz and Zatorre, 2012).

Not surprisingly, musical training is related to improvement in music-related task performance such as pitch discrimination (e.g., Kishon-Rabin et al. 2001; Micheyl et al. 2006; Magne et al., 2006; Bianchi et al., 2016; Zarate et al., 2012; Hou et al., 2014; Schön et al., 2004), melody recognition (e.g., Halpern et al., 1995; Herff and Czernochowski, 2017), timing perception (e.g., Rammsayer and Altenmüller, 2006; Repp, 2010; Matthews et al., 2016; Jones and Yee, 1997; Danielsen et al., 2022) and production (e.g., Cameron and Grahn, 2014; Danielsen et al., 2019; Fujii et al., 2011; Krause et al., 2010; Manning and Schutz, 2015; Manning et al., 2017, 2020; Repp and Doggett, 2007; Skaansar et al., 2019) or musical structure learning (e.g., Schön and François, 2011; François and Schön, 2011).

Yet, musical training has also been related to improvement of higher-order functions such as WM, IC and attention. Consequently, taking into account that WM, IC and attention are related to ASA, musical expertise may be an important factor potentially modulating it.

INFLUENCE OF MUSICAL EXPERTISE ON EXECUTIVE FUNCTIONS AND ATTENTION

The fact that musical training may improve executive functions, and particularly WM and IC, is not shocking considering that musicians must exert precise control over their own movements when playing their instrument, reproduce complex patterns of notes, pay attention to timing and pitch details and inhibit other sounds in their environment. However, such a transfer from a specific training to more general-domain functions is still a matter of debate. Firstly, because it is difficult to know whether musical training enhances executive functions or whether individuals with enhanced executive functions tend to become musicians (Schellenberg and Peretz, 2008). Secondly, because in reality,

mixed results can be found in the literature about the impact of musical expertise on executive functions (Benz et al., 2016).

WORKING MEMORY

That said, an important body of literature has investigated the influence of musical training on WM and some studies found better performance of musicians compared to non-musicians when musical stimuli were used. For instance, Pallesen and collaborators (2010) demonstrated that musicians outperformed non-musicians in a one-back and in a two-back task requiring memorization of chords. They also showed increased activity in brain regions thought to be involved in WM (e.g., prefrontal cortex, posterior parietal cortex, middle frontal gyrus) for musicians. Consistently, Ding and colleagues (2018) reported that musicians outperformed non-musicians in an N-back tonal task but not in a forward tone sequence discrimination task. Schulze and colleagues (2011a) found better performances for musicians than non-musicians in a task where participants were presented with five tones sequences followed by a probe tones, after 4 to 6 seconds, and asked to indicate whether this probe tone was in the sequence or not. In a subsequent work, Schulze and collaborators (2012) found that musicians keep tone sequences more efficiently in memory than non-musicians because of their better knowledge of musical structure. However, this finding was only true in a forward task, engaging no manipulation, and not in a backward task, suggesting that musicians have a better STM rather than a better WM. And this result is consistent with other studies showing increased STM in musicians (e.g., George and Coch, 2011; Pallesen et al., 2010; Schulze et al., 2011b; Monahan et al., 1897).

Other studies, employing visual and verbal material, also found improvement of musicians' WM. Bugos and colleagues (2007) for example, reported a better WM for musicians using verbal material. George and Coch (2011) found similar results using both forward and backward digit span tasks, as well as in a visual spatial WM task. They also found an earlier P300 ERP in musicians suggesting faster updating in WM (Donchin and Coles, 1988; Polich, 2007). In addition, they reported a larger P300 amplitude for musicians, suggesting that they allocated more neural resources to auditory stimuli. Consistently, several other studies found a musician's advantage using the backward digit span task (Clayton et al., 2016; Suarez et al. 2016; Zuk et al., 2014; Parbery-Clark et al., 2009a, 2011; Grassi et al., 2017; Talamini et al., 2016). Finally, Franklin and collaborators

(2008) revealed a musician's advantage in memorizing words while simultaneously carrying arithmetic operations (see also Lee et al., 2007).

However, some studies fail at finding an effect of musical expertise on WM performance using visuo-spatial forward and a backward span task (Bialystock and DePape, 2009), backward digit span (Hansen et al., 2012; Schulze et al., 2012; Boebinger et al., 2015) as well as composite WM scores including arithmetic WM task, forward, sequencing and backward digit span task (Talamini et al., 2022).

At last, Talamini and colleagues (2017) revealed that across 16 studies and 19 tasks, the better performance of musicians compared to non-musicians in WM tasks was consistent when tonal material was employed, with a large effect-size ($n = 3$, $g = 1.04$, $CI = .48 - 1.6$) as well as when verbal material was used, with a small to moderate effect size ($n = 13$, $g = .59$, $CI = .34 - .84$). However, it was not the case for visuo-spatial material ($n = 3$, $g = 0.1$, $CI = -.5 - .52$).

Overall, these results suggest a better working memory for musicians than non-musicians, especially with tonal and verbal material, even if some studies failed to detect such a difference.

INHIBITORY CONTROL

Akin to WM, several studies investigated the link between musical expertise and IC. And once again, results are ambivalent.

For instance, Moreno and colleagues (2011) found enhanced performance in a Go/NoGo task between two test sessions for children who followed a musical training compared to children who followed an art training. In a subsequent study, Moreno and collaborators (2014) found no difference between musicians and non-musicians in terms of performance in a visual Go/NoGo task but they did find difference in terms of ERP waveforms. Musicians showed enhanced early P2 and reduced N2 amplitude. The authors proposed that this pattern could reflect the fact that musicians may experience a lesser degree of response conflict and therefore, require less response inhibition than non-musicians. Later, Moussard and colleagues (2016) reported similar results meaning, no behavioral difference in a Go/NoGo task but differences in ERP waveform between musicians and non-musicians. Bialystok and DePape (2009) found that musicians

outperformed non musicians in a visual Simon task as well as in an auditory Stroop task but only for overall response time. Very recently, Sharma and colleagues (2019) reported faster RT and lower error rate for musicians than non-musicians in classical auditory Stroop tasks using pitch names (“low” and “high”) but also in auditory Stroop tasks employing either their English solmization or their phonemic equivalents. However, they only measured RT and error rate separately, without taking into account the Speed-Accuracy Trade-Off. Meyer and colleagues (2018) also found better performance of musicians compared to non-musicians in a visual Flanker task. That being said, interference effect in the flanker task is thought to be mainly due to *stimulus-response conflict* (Eriksen and Eriksen, 1974; Eriksen and St James, 1986; Eriksen, 1995; van Veen et al., 2001).

More ambivalent are the results of Kaganovich and colleagues (2013) who found that musicians outperformed non musicians in a task requiring inhibition of distracting task-irrelevant changes (the *auditory distraction paradigm*, Schröger and Wolff, 2000). This effect was important when the deviant was a musical sound but it disappeared when the deviant sound was a voice which suggested a better ability to resist distraction for musicians when the distracting stimulus remains their domain of expertise. Consistent with these results, Schellenberg (2011) found no difference in attentional inhibition measure with an adapted version of the Stroop (Moon Stroop) in children with or without musical training. Similarly, Vanden Bosch der Nederlanden and collaborators (2020) found no difference in the magnitude of the Stroop effect in a Color-Words Stroop task between musicians and non-musicians (see also Clayton et al., 2016 and D’Souza et al., 2018). Finally, in an fMRI study, Sachs and colleagues (2017) reported no difference in the performance of a classical Stroop task between musically trained children and non-musicians children despite greater activation in regions involved in cognitive control. Interestingly, they also reported no difference between children involved in musical training and children involved in sport training in both behavioral measures and BOLD responses, suggesting that extracurricular training may promote structural brain change even if it is not a musical training.

Overall, these mixed findings seem to suggest that musicians could be better than non-musicians in inhibitory control but that this advantage could only be present in response inhibition and for specific material (musical sounds). This is not surprising considering

that musicians probably spend more time exerting response inhibition rather than attentional inhibition when they are playing their instrument. Indeed, musical instruments are generally played with both hands requiring uncoupled movements. Sometimes with different rhythms (e.g., drums) and/or different motor activity (e.g., violin). Thus, learning to play an instrument may imply exerting response inhibition to be able to decouple hand's movements. In addition, playing an instrument, especially in an orchestra, requires fine rhythm synchronization which involves response inhibition.

ATTENTION

On the margins of executive function *per se*, the research community has also investigated the relations between musical expertise and attention. Indeed, in a very recent work, Medina and Barraza (2019) reported a better efficiency of executive control network which, as defined by Posner and Petersen (2012), is involved in the suppression of distracting stimuli (response inhibition) and in the top-down attentional control (attentional inhibition). However, they used the Attentional Network Test (Fan et al., 2002; Posner and Petersen, 1990) which is a modified version of the Flanker task. In addition, they did not take the Speed-Accuracy Trade-Off into account in their interference score. Thus, it is unclear whether the musicians' advantage in executive networks reported here is due to a better response inhibition or a better attentional inhibition. Likewise, Roman-Caballero and colleagues (2021) used a slightly modified version of the ANT to investigate the impact of musical expertise over the alerting, orienting and executive control networks of attention (Posner and Petersen, 2012). They found a better efficiency of the alerting network for musicians compared to non-musicians but no difference for the orienting and executive control network which is inconsistent with the work of Medina and Barraza (2019). In addition, they reported reduced executive vigilance decrement over time-on-task for musicians compared to non-musicians indicating a possible better sustained attention. In another work, Strait and colleagues (2010), using an auditory Go/NoGo task, reported faster reaction times to target cues for musicians compared to non-musicians that they interpret as an enhancement of auditory attention and a possible reduced variability in musicians sustained auditory attention. Conversely, they did not find any difference in an analog visual task (see also Roden et al., 2014 and Clayton et al., 2016). These results were subsequently replicated by Strait and Kraus (2011). A later work from the same team

(Strait et al., 2014) prolonged the previous studies using a developmental approach. They investigated the impact of auditory expertise during early childhood on the cortical index of selective auditory attention. They found that selective attention-related prefrontal evoked responses became less variable with musical expertise suggesting better selective attention and inhibitory control in adult and school-age musician children than for their non-musicians peers. In a fMRI study, Pallesen and colleagues (2010) found larger BOLD responses for musicians in several regions supposed to be involved in both sustained attention and cognitive control (e.g., lateral prefrontal cortex, insula, posterior dorsal prefrontal cortex) during a task in which participants were listening to musical material (tones, timbre). Which is consistent with other studies showing greater activation for musicians than non-musicians, in cortical areas engaged in sustained auditory attention and working memory (e.g., Strait and Kraus, 2011; Baumann et al., 2007, 2008; Gaab and Schlaug, 2003; Haslinger et al., 2005).

Overall, it seems that attention could be enhanced in musicians compared to non-musicians. Once again, it seems that musical training may specifically improve the alerting system leading to an increased sustained attention ability due to a better monitoring of arousal level (Roman-Caballero et al., 2021). Although evidence for enhanced selective attention have also been provided, they seem to be less consistent.

That being said, because attention is supposed to be dependent on WM and IC, as shown by overlapping brain regions, it is difficult to know whether musicians really have better attentional capacities or whether they have better IC and/or WM leading to a better sustained and/or selective attention. Further work is required to disentangle this node.

EVIDENCE OF THE INFLUENCE OF MUSICAL EXPERTISE ON AUDITORY SCENE ANALYSIS

Beside the effects of musical practice on EF *per se*, several works have investigated its influence on ASA. But once again, results are equivocal.

For instance, Alain and Zendel (2009) found that musicians were better at segregating complex musical sounds that occur simultaneously based on harmonicity than non-musicians which is in line with previous similar works (Munte et al., 2001; Nager et al., 2003; Oxenham et al., 2003; van Zuijen et al., 2004). Other studies reported better performance of musicians in clinical speech-in-noise tests (QuickSIN and HINT; Parbery-

Clark et al., 2009a) and in tasks requiring participants to repeat words presented in multitalker babble noise (Zendel and Alain, 2015). Using a protocol that was more representative of a cocktail-party situation, Swaminathan and colleagues (2015) reported better performance of musicians in a task requiring participants to listen and understand a talker while ignoring concurrent spatially separated intelligible speech streams. Consistently, Clayton and collaborators (2016) found better performance of musicians compared to non-musicians in a spatial hearing task where participants need to identify a target word embedded in a sentence, among masker sentences. Finally, very recently, Bidelman and Yoo (2020) extended these results and reported better performance of musicians compared to non-musicians in a complex realistic multitalker cocktail-party scenario in which they used up to eight concurrent talkers and where the target speech stream as well as the masker ones were located in the entire 360° soundfield.

Using EEG, Parbery-Clark and collaborators (2009b) showed that this behavioral musician's advantage goes along with a more robust neural response (cortical and subcortical) to speech sounds. This result is supported by other studies showing that musicians have more robust frequency-following responses (Bidelman and Krishnan, 2010; Musacchia et al., 2008), enhanced auditory evoked potential (Shahin et al., 2003) and event related potentials (Baumann et al., 2008; Zendel and Alain, 2009, 2013) than non-musicians. In addition, it is reinforced by works demonstrating enhanced cortical representation of piano tones (Pantev et al., 1998, Kuriki et al., 2006), enhanced entrainment of slow neuronal oscillations (1-8Hz) to the dominant note rate of musical stimuli (Drilling and Poeppel, 2015) as well as enhanced cortical tracking of synthetic musical (Di liberto et al., 2020b) and speech stimuli (Puschmann et al., 2019).

Nevertheless, several studies still failed to find musician's advantage in a variety of conditions including clinical speech-in-noise tests (Ruggles et al., 2014), tasks requiring to understand sentences in either continuous or gated noise (Ruggles et al., 2014), or in the presence of maskers (Boebinger et al., 2015). Interestingly, using a rather ecological spherical array of 64 loudspeakers, Madsen and colleagues (2019) found no better performance of musicians compared to non-musicians in understanding speech in noise under a variety of conditions including difference in spatial separation, speech material and reverberation.

Consequently, it seems that the representation of speech and music sounds is enhanced in musicians at different stages including brainstem and cortical areas, compared to non-musicians. However, despite the assumed enhanced subcortical and cortical sound representation of musicians, several studies also failed at finding behavioral differences in complex situations between musicians and non-musicians.

In the light of these results and as suggested by Bidelman and Yoo (2020), musicians may be better able to understand speech in presence of noise or concurrent talkers but this advantage may only occur in very challenging conditions. And this advantage could arise from the enhanced sound representation in challenging conditions at early stages of the auditory processing (e.g., Bidelman and Krishnan, 2010; Parbery-Clark et al., 2009b, 2011).

SUMMARY

In the light of the previous sections, it is not straightforward to claim that musicians are better in ASA than non-musicians. Indeed, on the one hand, there are cues in favor of better WC, IC and attention in musicians (i.e., important functions underlying ASA), better performance in auditory streaming tasks in very challenging situations and better cortical and subcortical sounds representations. However, on the other hand, several studies failed at finding behavioral evidence supporting this musician's advantage in the cognitive functions underlying ASA as well as in complex listening situations.

All these elements may suggest that musicians could benefit from their training only in specific situations that looks like their domain of expertise and therefore require similar brain processes (e.g., pitch discrimination, fine temporal analysis, inhibition of response). Consequently, the question of a far-transfer of musical expertise on general auditory streaming ability remains unanswered.

AUDITORY ATTENTION DETECTION

The major part of this chapter is taken from the article “EEG-based auditory attention detection and its possible future applications for passive BCI” from Belo, Clerc and Schön, published in 2021 in Frontiers in Computer Science.

OVERVIEW

This chapter is dedicated to explain the framework of Auditory Attention Detection (AAD). First, the concept of electroencephalography is briefly reviewed as it is necessary to understand how AAD was born. Then, the physiological foundations of AAD are exposed. Afterward, the different algorithms and philosophies of the technique are presented as well as the main performance metrics and the global limitations.

WHAT IS ELECTROENCEPHALOGRAPHY?

Basically, electroencephalography (EEG) is a brain imagery technique that allows one to measure the summed electrical activity generated by vast populations of neurons (Lopes da Silva, 2010). Indeed, neurons produce transmembrane electric current when activated. Although one can distinguish several kinds of neural activation (Lopes da Silva and van Rotterdam, 2005; Lopes da Silva, 2002) and neurons types, the take home message here is that because neurons are interconnected and arranged in columns in the cortex perpendicularly oriented relative to the scalp (Mountcastle, 1997), their synchronized activation produce enough electrical activity to be recorded by electrodes located outside the brain. The main problem being that it is very complicated, if not impossible, to localize the source of the electrical potential measure with an electrode because neurons are interconnected and also because of the tissues (scalp, skull, cerebrospinal fluid layer, brain) conductivity. This is what is called the inverse problem of EEG (von Helmholtz, 2004). Consequently, tissue conductivity induces inter-individual variability that is difficult to control for and independent of the cognitive activity measured with the system. In addition, one can say that spatial resolution of EEG is quite poor, except for single unit measurement using intracranial EEG (see section below). But on the other hand, temporal resolution is extremely good (in a sub millisecond order) and only limited by the sampling frequency of the amplifier which can reach 30 kHz in certain cases (Lachaux et al., 2003).

Materially speaking, an EEG system is composed of a set of electrodes and an amplifier that will amplify the measured electric potential, since a population of around 50 000 neurons would generate an electric field with a magnitude of no more than 10 nA (Murakami et al., 2006). In reality, there are several types of EEG. The first distinction that can be made is between extracranial and intracranial EEG. In the intracranial case, electrodes are directly implanted inside the brain. In general, the utilization of intracranial EEG is restricted to specific populations such as pharmaco-resistant epileptic patients, Parkinson patients or patients with a tumor. In the case of epilepsy for instance, intracranial EEG is used to precisely localize the epileptic focus for the purpose of surgically removing it. Intracranial electrodes are of different types. They can be arrays of electrodes that could be implanted in deeper structures of the brain such as the thalamus or they can be grids of electrodes that will be arranged on the top of the cortex. Temporal as well as spatial precision is very good with intracranial EEG because it does not suffer from the impact of tissue conductivity. True to its name, scalp EEG (or extracranial EEG) refers to a system where electrodes are arranged on the outer part of the head (the scalp). In the case of scalp EEG, electrodes are located on a cap which contains more or less electrodes (from 8 to 256) and this cap is positioned on the head of the user (either a patient or a participant). Because of the above-mentioned impact of tissue conductivity on the electric current, scalp EEG spatial precision is low. However, its temporal precision is still very good (in the order of a millisecond) and widely enough to tackle cognitive processes that generally occur after a few milliseconds. Another distinction that can be made concerns the type of electrode. Indeed, electrodes can be wet or dry and active or passive (Mathewson et al., 2017). The difference between wet and dry electrodes lies in the fact that wet electrodes utilize electrolyte gel to improve the contact with the skin and consequently reduce impedance while dry electrodes only use mechanical force (Taheri et al., 1994). While wet electrodes show better signal to noise ratio than dry ones, they are less suitable for real-life or out-of-the-lab experiments. Concerning passive and active electrodes, the latter are supposed to be less sensitive to environmental noise (Kappenman and Luck, 2010; Laszlo et al., 2014) because each electrode is amplified directly at the scalp source (Metting Van Rijn et al., 1996) which is interesting knowing that the quantity of non-neural noise impacts the statistical power of EEG recordings (Luck, 2014).

To summarize, EEG is a brain imagery technique that measures the electrical activity of million neurons with a very good temporal resolution, allowing the non-invasive measurement of cognitive processes. In addition, it has the advantage of being possibly adapted for out-of-the-lab conditions.

EEG-BASED AUDITORY ATTENTION DETECTION

BASICS OF EEG-BASED AUDITORY ATTENTION DETECTION

EEG-based auditory attention detection (AAD) emerged in the beginning of the 2010's. This approach relies on two fundamental properties of the brain. The first one is that the amplitude envelope of a dynamic acoustic stimulus, such as speech or music, is represented at a cortical level (Mesgarani et al., 2009; Nourski et al., 2009). More precisely, in the theta-delta and the gamma oscillatory activity in the human auditory cortex (Giraud and Poeppel, 2012; Pasley et al., 2012; Kubanek et al., 2013, Doelling and Poeppel, 2015). The second one is that paying attention to an acoustic source result in a greater coupling between the amplitude envelope of the source and the amplitude envelope of neural activity in these bands, especially in a dichotic context where the cortical tracking of the attended source is enhanced compared to the unattended one (Zion Golumbic et al., 2013). These two important findings have led to works that demonstrating the possibility to detect which auditory stream, within a set of multiple concurrent streams, an individual was attending to using neural data recorded via electroencephalography (EEG) or magnetoencephalography (MEG) (e.g., O'Sullivan et al., 2015, 2017; Akram et al., 2014; Mirkovic et al., 2015; Zink et al., 2016, 2017; Fiedler et al., 2017; de Taillez et al., 2017).

This new research domain has grown rapidly in the last decade and now a lot of different algorithms to detect auditory attention are available.

TWO APPROACHES: LINEAR AND NONLINEAR

One can distinguish two main categories of approaches to detect auditory attention: linear and non-linear models (see Geirnaert et al., 2020 for a comprehensive review of AAD Algorithms).

LINEAR MODELS

In the community of linear models, two main “philosophies” are in competition (see Alickovic et al., 2019 for a complete review on linear models): forward, or encoding (encoding because these models are a description of how the system encodes information), and backward, or decoding, models. The objective of the forward strategy is to predict the neural response in the neural data (i.e., EEG channels) from the representation of the audio signal via a temporal response function (TRF, i.e. an encoder) that describes the linear relationship between a set of neural data and an audio stimulus at certain time points (Crosse et al., 2016). In the simplest case (i.e., one audio signal) a unique representation of the audio signal is created. This representation can be the amplitude envelope (O’Sullivan et al., 2015), the spectrogram of speech signal (O’Sullivan et al., 2017), or the Mel spectrogram for a music signal (Cantisani et al., 2019). Depending on the type of the chosen representation the analysis can be either univariate (an amplitude envelope is a univariate stimulus feature) or multivariate (a spectrogram is a multivariate stimulus feature). Although it is possible to use multivariate TRF with the forward approach, this strategy is, by nature, univariate (Crosse et al., 2016). Afterward, the audio representation is convolved with an unknown channel-specific TRF. To estimate the TRF (i.e., fit the model parameters), an error minimization is performed between the neural response and the one predicted by the convolution (e.g., Mean-Squared Error) using assumptions about noise distribution (Holdgraf et al., 2017). Once the model’s parameters have been estimated, the model is validated on new data. These new data could be from the same dataset used to estimate the parameters (leave-n-out procedure) or from data recorded separately. The validation step is crucial because, to be interpretable, the model should be compatible with new data and make accurate predictions (generalization ability). Finally, the rationale of the forward strategy, in auditory research, is to predict neural data on the basis of the sound’s features.

Backward models work similarly but by predicting the auditory representation based on neural data (Alickovic et al., 2019). A pre-trained neural linear decoder is applied to the neural data to reconstruct the chosen representation (this is the reason why this type of approach is sometimes called “stimulus reconstruction”). The reconstructed representation is compared to the original representations. A high similarity (correlation) indicates a good performance of the model. Two other approaches can also

be mentioned: Canonical Correlation Analysis (CCA) and Bayesian state-space modeling. Canonical Correlation Analysis is a hybrid model that combines a decoding and an encoding model. This approach, developed by de Cheveigné and colleagues (2018), aims to minimize the irrelevant variance in both neural data and stimulus by a linear transformation. Concerning Bayesian state-space modeling (Miran et al., 2018b), it is composed of three modules: a dynamic encoder/decoder estimation module, an attention marker extraction module, and a real-time state-space estimator module (see Miran et al., 2018a for a complete description of the model) and this approach was developed in the purpose of real-time decoding of auditory attention. As mentioned before, in the context of AAD, linear models are generally used with two (or more) concurrent speech streams in order to determine which stream the listener is attending to. In this case, a representation of each auditory source is created (e.g., speaker 1 and speaker 2). Once the model has been fitted, no matter which approach was chosen, a two-class classifier is used to decide which of the two streams the participant was focused on. To do so, the classifier compares the correlation coefficients between the model output and the original model input representations (e.g., the correlation between the reconstructed envelope and the original audio signals envelopes in backward strategy) over a certain portion of data (decision time windows). The highest correlation indicates which stream the participant was attending to. The length of the decision time window is a crucial parameter because correlation-based measures need a certain amount of information to perform well. However, short decision time windows (<2 s of data) are of interest in Brain Computer Interface for real-time classification.

NONLINEAR MODELS

Similarly to linear models, several non-linear model architectures are in competition. But non-linear models are still overlooked because they are more complex to implement and interpret. Nevertheless, they were used by a few studies to explore AAD. Vandecapelle and colleagues (2020) used two convolutional neural networks to determine the attended speaker in a multi-speaker scene by using the direction of the locus of auditory attention. Their method allows them to decode auditory attention with very short decision time windows and with a good classification accuracy (around 80% for 2 s of data). In another study the authors used a fully-connected neural network to reconstruct the speech envelope and estimate the attended speaker (de Taillez et al., 2017). The

classification accuracy obtained with this method appears to be similar to the performance obtained in Vandecapelle et al. (2020) even though the comparison between studies is not straightforward due to differences in experimental and model parameters or accuracy measures (Ciccarelli et al., 2019). However, non-linear models outperform linear models in terms of decision time window/performance ratio. One other potential advantage of this type of model is that it seems more realistic insofar as it may capture the neuronal non-linearity underlying speech perception (O’Sullivan et al., 2015, Mirkovic et al., 2015, de Taillez et al., 2017).

AAD PERFORMANCE

Generally, AAD performances are assessed with two accuracy metrics: regression accuracy and classification accuracy (Wong et al., 2018). Regression accuracy evaluates the goodness of fit of the model and it is expressed in terms of correlation coefficient (Pearson’s correlation, often ranging 0.05–0.2) between the output of the model and the real value (e.g., speech envelope is correlated with reconstructed envelope for backward models). Classification accuracy, on the other hand, evaluates the ability of the classifier to correctly identify the attended stream for a given decision time window and it is generally expressed in terms of percentage of good classification. Classification accuracy is generally high for long decision time windows (around 85% for 60 s of data) but drops drastically for shorter decision time windows no matter which approach is used. Recently, Wong et al. (2018) showed that decoding models outperform encoding models in terms of classification accuracy. One of the best classification results obtained so far was 85% with 20-s decision time windows, with the CCA (Geirnaert et al., 2020).

Note that, very recently, Geirnaert and colleagues (2020) proposed a new metric, named MESD (for *Minimal Expected Switch Duration*), to estimate the performance of an AAD algorithm that takes into account both accuracy and decision window size length. Besides allowing the comparison of performance between different algorithms and different studies (which is always difficult because performance is not always described in the same terms across different studies), this metric is also useful in the framework of neurosteered hearing aids.

LIMITATIONS

Linear and non-linear models still suffer from several limitations with respect to AAD. The major problem of linear models lies in the fact that their classification accuracy is strongly influenced by the duration of the decision window. Long windows yield good classification (>80%) while short ones (e.g., 2 s) yield much poorer performance (~60%). This is due to the fact that (1) short decision windows contain less information (Vandecapelle et al., 2020), (2) EEG signals contain a mixture of several physiological and neural processes. Thus, correlations between predicted and actual data are rather weak (between 0.05 and 0.2) and short decision time windows are particularly sensitive to noise (Geirnaert et al., 2020). Moreover, a huge amount of data is needed to fit the model properly. Therefore, these models are difficult to use in real time situations where the selection of the attended speaker must be performed as fast as possible. For non-linear models, the principal issue is the risk of overfitting, in particular with small datasets (Vandecapelle et al., 2020). Moreover, comparing performances of several non-linear models on different datasets pointed to a low reproducibility of these algorithms (Geirnaert et al., 2020).

Besides fitting issues, physiological noise and non-relevant neural signal, another source of performance variability could reside in inter-individual differences at the cognitive level (Ciccarelli et al., 2019; Jaeger et al., 2020), such as for instance in working memory, inhibitory control, selective and sustained attention but also motivation and fatigue. However, until now, no work has been done to explore the link between cognition and AAD performance.

RATIONALE BEHIND THIS WORK

An important part of the cited works in the former sections of this manuscript used speech stimuli. And this is particularly true for the studies exploring AAD. To our knowledge, only a few studies used music stimuli in this framework (e.g., Di Liberto et al., 2020b; Cantisani et al., 2019; Madsen et al. 2019; Hausfeld et al., 2021). However, in the study of Di Liberto and colleagues (2020b) synthetic versions of musical pieces were used which remove the ecological dimension. Madsen and colleagues (2019) used only monodic stimuli which did not allow them to investigate ASA in complex situations. Finally, in the work of Cantisani and collaborators (2019) they used real recording of

monophonic instruments but they apparently did not control for their saliency. It is thus possible that some of their stimuli were easier to follow because they captured attention to a greater extent. Therefore, an objective of this work is to demonstrate the possibility of reconstructing highly ecological polyphonic musical excerpts, that are matched in rhythm, timbre and genre, from EEG-recorded neural data using a AAD stimulus reconstruction approach and investigating whether neural data allow detecting the focus of attention in a dichotic music listening context.

As mentioned in the previous chapter, AAD performance is highly variable from one individual to another. And for the moment, it is still unknown why we observe such a variability. One possible explanation could reside in inter-individual differences at the cognitive level. More specifically, AAD performance, in dichotic context, is mainly the result of an ASA which is supported by the ability to selectively focus attention on a specific sound source while ignoring the concurrent one. Yet, this capacity is supported by several high-order functions such as attention, WM and IC, as presented in the former chapter of this manuscript. However, no work has been carried on yet, to explore the link between AAD performance and these specific cognitive functions. Consequently, the most important goal of this work is to estimate to what extent sustained attention, WM and attentional inhibition can explain the inter-individual variability of AAD in terms of reconstruction and classification performances of polyphonic musical stimuli. We hypothesize that a better sustained attention as well as a better WM capacity and a better attentional inhibition will lead to a higher aptitude at focusing attention to a target stimulus and ignoring a concurrent stimulus presented simultaneously in a different ear, which will result in higher AAD performance.

A third objective of this work is to explore the impact of musical expertise on both cognitive functions (sustained attention, WM and IC) and AAD performance. Indeed, as mentioned earlier in this manuscript, although it is still controversial, musicians are thought to have better cortical tracking of sounds as well as better higher order functions and attentional functioning. They should, therefore, have better AAD performance than non-musicians. Consequently, we hypothesize that musicians will have better sustained attention, WM and IC scores than non-musicians as well as better AAD performances in monodic and dichotic conditions because of their superior cognitive functions.

That being said, we saw in the former chapters that mind wandering as well as prior knowledge can also affect sound representations. Therefore, these factors can have an influence on the AAD reconstruction accuracy. Consequently, the last objective of this thesis was to investigate to what extent the familiarity and the mind wandering, measured via questionnaires, can explain AAD reconstruction performance of monodic trials. We hypothesize that familiarity will have a positive effect over reconstruction accuracy while mind wandering will have a negative effect

EXPERIMENTAL WORK

STUDY ONE: ATTENTIONAL INHIBITION ABILITY PREDICTS NEURAL REPRESENTATION DURING CHALLENGING AUDITORY STREAMING

The major part of this chapter is taken from the article “Attentional inhibition ability predicts neural representation during challenging auditory streaming” from Belo, Clerc and Schön, submitted in 2022 in The Journal of Cognitive Neuroscience.

INTRODUCTION

The human ability to concentrate attention on one specific sound source in a complex auditory environment is a fundamental skill that allows efficient communication. Indeed, it allows us to follow conversations in crowded places such as classrooms, family dinners or parties. Even if the auditory system seems to be able to perform this task almost effortlessly, concentrating attention on one sound source remains a challenging task that requires a variety of cognitive processes. Such a challenge is magnified for people suffering from profound hearing loss. This is mainly due to the fact that cochlear implants (CI) suffer from a poor spectral resolution (Zeng et al., 2008) and do not allow selective amplification of the source of interest, yielding to a rather poor auditory perception in a noisy environment or in presence of simultaneous auditory sources (Moore, 2003; Crandell, 1991; Humes et al., 1996). Moreover, it has been hypothesized that the early auditory deprivation may lead to cognitive dysfunctions, also contributing to a poor speech-in-noise comprehension and auditory scene analysis. Therefore, it is particularly important to find a way to either improve auditory scene analysis abilities in CI users or more directly, to improve CI’s amplification systems.

To address this issue, a new approach, called M/EEG-based auditory attention detection (AAD), emerged a few years ago. This approach is based on several works that have demonstrated the possibility to detect which auditory stream, within a set of multiple concurrent streams, an individual was attending to using neural data recorded via electroencephalography (EEG) or magnetoencephalography (MEG) (Ding and Simon, 2012; O’Sullivan et al., 2015; Akram et al., 2014). This relies on the fact that the brain tracks the amplitude envelope of auditory stimuli at the cortical level (Mesgarini et al.,

2009; Ding and Simon, 2012, Giraud and Poeppel, 2012; Nourski et al., 2009; Kubanek et al., 2013). Moreover, in a dichotic context (i.e., different sound in each ear), the cortical tracking of the attended source is enhanced compared to the unattended source (O'Sullivan et al., 2015; Pasley et al., 2012; Zion-Golumbic et al., 2013). Therefore, these findings are of particular interest to develop neuro-steered hearing aids that are able to detect the source of interest and subsequently amplify it.

However, the literature shows a rather high inter-individual variability in the AAD performance. This variability can be partly explained by physiological factors such as the cranial thickness, the cortical orientation with respect to the skull or the signal to noise ratio of neural data. Nevertheless, it could also be explained by behavioral and cognitive factors including stimulus familiarity, attention, motivation, fatigue and executive functions (Ciccarelli et al., 2019).

Undoubtedly, the ability to selectively listen to a source in a multi-source environment relies upon several cognitive functions. At the outset, focusing attention on one source requires effective segregation of the distinct sources in the auditory scene implying both bottom-up and top-down mechanisms (Bregman, 1994; Kondo et al., 2017). The formation of sound streams is particularly dependent on attention because it tends to force the auditory system to group sound elements in line with behavioral or perceptual goals (Shamma et al., 2011; Sussman, 2017; Loui and Wessel, 2007; Snyder et al., 2006). Beside stream segregation per se, it has been shown that attention is important for speech in noise comprehension (Wild et al., 2012). But other, yet related, executive functions seem to be crucial in tasks that require maintaining attentional focus on a source in a cocktail-party scenario.

Firstly, the process of maintaining attention on a task or an auditory object during a certain amount of time is a critical skill for auditory scene analysis and partly depends on sustained attention (Gadea and Espert, 2009; Esterman and Rothlein, 2019; Tierney et al., 2019). Indeed, sustained attention abilities are linked to better speech in speech comprehension (Tierney et al., 2019) and are also involved in dichotic listening (Helge Johnsen et al., 2002; Hommet et al., 2010; Asbjornsen and Hugdahl, 1995; Hugdahl and Andersson, 1986). However, sustained attention is closely related to other functions and concepts such as selective attention, working memory, inhibitory control, vigilance or

motivation (Esterman and Rothlein, 2019). Therefore, it seems to be more a construct than a single cognitive function.

Secondly, working memory (WM) is also important because it is related to speech in noise perception (Besser et al., 2013; Sörqvist and Ronnberg, 2012; Souza et al., 2015; Escobar et al., 2020), dichotic listening (Conway et al., 2001; Colflesh and Conway, 2007; James et al., 2014) and in polyphonic music listening (Janata et al., 2002). Additionally, it also seems to play a role in the segregation of streams (Goll et al., 2012) and in the representation of auditory objects (Christison-Lagay et al., 2015) probably because these processes are achieved in the temporal domain. Finally, WM is related to involuntary attentional switching in high cognitive load conditions (Berti and Schröger, 2003).

Finally, inhibitory control seems to be essential to selectively listen to a source in a multi-source environment because inhibition of concurrent streams is necessary. Actually, attentional as well as response inhibition have been shown to be involved in dichotic listening task (Hugdhal et al., 2009; Falkenberg et al., 2011), speech in noise (Dryden et al., 2017; Knight and Heinrich, 2017; Campbell et al., 2020) or auditory scene analysis (Goll et al., 2012). Additionally, in cocktail-party-like scenarios, the different sources can act as distractors and capture attention. Not being able to resist distractions can then lead to involuntary attentional switches resulting in the loss of the source of interest. Moreover, when the cognitive load is high, re-focusing on a target source can be effortful or even impossible, due to the limited amount of attentional resources available (Navon, 1984).

In light of all these elements, WM, sustained attention as well as attentional inhibition could impact performances of AAD algorithms because they support the cognitive processes required when listening to complex auditory scenes.

Yet, another factor could influence the performance of AAD. Actually, in a very recent study, using monodic synthetic musical stimuli, Di Liberto and colleagues (2020b) demonstrated that average reconstruction performances of their AAD were higher for musical experts than participants with no musical expertise. This result corroborates previous findings showing a better cortical tracking of auditory stimuli for musicians compared to non-musicians (Doelling and Poeppel, 2015; Puschmann et al., 2019). Musicians are known to be better in music related tasks such as pitch discrimination

(Kishon-Rabin et al., 2001; Micheyl et al., 2006; Magne et al., 2006; Bianchi et al., 2016) or timing perception (Rammsayer and Altenmüller, 2006). They are also better in stream segregation (Zendel and Alain, 2009; François et al., 2014), speech in noise listening (Parbery-Clark et al., 2009a; Coffey et al., 2017) and tend to have better performances in dichotic listening (Barbosa-Luiz et al., 2021). In addition, some works have shown better scores in musicians than in non-musicians for cognitive functions involved in cocktail-party like tasks, including cognitive control (Bialystok et al., 2009; Moreno and Farzan, 2015; Moussard et al., 2016; Kaganovich et al., 2013), WM (George and Coch, 2011; Jakobson et al., 2008; Zuk et al., 2014; see Talami et al., 2017 for a review) and attention (Medina and Barraza, 2019; Strait et al., 2010, 2015; Strait and Kraus, 2011; Clayton et al., 2016).

The first objective of this study was to demonstrate the possibility of reconstructing highly ecological polyphonic musical excerpts from EEG-recorded neural data using a stimulus reconstruction approach and investigating whether neural data allow detecting the focus of attention in a dichotic music listening context. The second and most important goal was to estimate to what extent sustained attention, WM and attentional inhibition can explain the inter-individual variability of AAD in terms of reconstruction and classification performances. Finally, the third objective was to explore whether musicians are better than non-musicians in terms of both monodic and dichotic AAD performances.

MATERIAL AND METHODS

PARTICIPANTS

A total of forty-five subjects took part in the experiment (28.93 ± 10.94 years, min = 18 years, max = 54 years, 22 females). This number was determined *a priori*, based on previous studies with the same approach (e.g., O'Sullivan et al., 2015; Di Liberto et al., 2020b; Cantisani et al., 2019). Over these forty-five participants, one did not complete the online cognitive tests and was thus excluded from the regression analysis, one did not terminate the listening task, and three were discarded from the analysis due to issues related to data acquisition.

All the participants included in the analysis reported no history of hearing disorders, attention impairment or neurological disorder and were not under medication.

Twenty-one participants reported having taken formal musical training and two declared being self-taught. Most musicians reported more than 15 years of music making (mean=16.08, sd=5.4) and daily practice. For eight of them, the piano was their main instrument. Four played the guitar and the others played the flute, french horn, harp, saxophone, trombone, trumpet or double-bass. For musicians the mean age was 34.6 ± 13.6 years and 11 of them were females while for non-musicians the mean age was 24.7 ± 6.6 years and 9 of them were females. The experiment was approved by the Comité Opérationnel d'Évaluation des Risques Légaux et Éthiques (COERLE) of INRIA and was undertaken in accordance with the Declaration of Helsinki. Each participant provided written informed consent and was financially compensated.

MAIN EXPERIMENTAL TASK

STIMULI

Monodic stimuli consisted of 60 second recordings of four well-known and four less-known classical instrumental piano pieces. The well-known musical excerpts were taken from Beethoven's Für Elise, Bach's Prelude in C Major, Mozart's Sonata in C Major K545 and Beethoven's Pathétique Sonata. The less-known musical excerpts were taken from Beethoven's Lustig und Traurig, Scarlatti Sonata in A Minor K54, Beethoven's Sonata in F Presto and Bach's Prelude no°8.

Dichotic stimuli consisted of pairs of musical excerpts simultaneously presented to the left and the right ear. Pairs were composed of rhythmically similar musical excerpts as assessed via Power Spectrum Density measures (see Figure 8). This allowed to ensure that the attention would not be systematically more captured by a greater presence of notes in one excerpt of the pair. The pairs were Beethoven's Fur Elise/Bach's Prelude in C Major, Mozart's Sonata in C Major K545/Beethoven's Pathétique Sonata, Beethoven's Lustig und Traurig/Scarlatti's Sonata in A Minor K54 and Beethoven's Sonata in F Presto /Bach's Prelude no°8. All musical excerpts were RMS normalized at the same amplitude.

Stimuli were presented via Sennheiser HD-25 supra-aural headphones at a comfortable level.

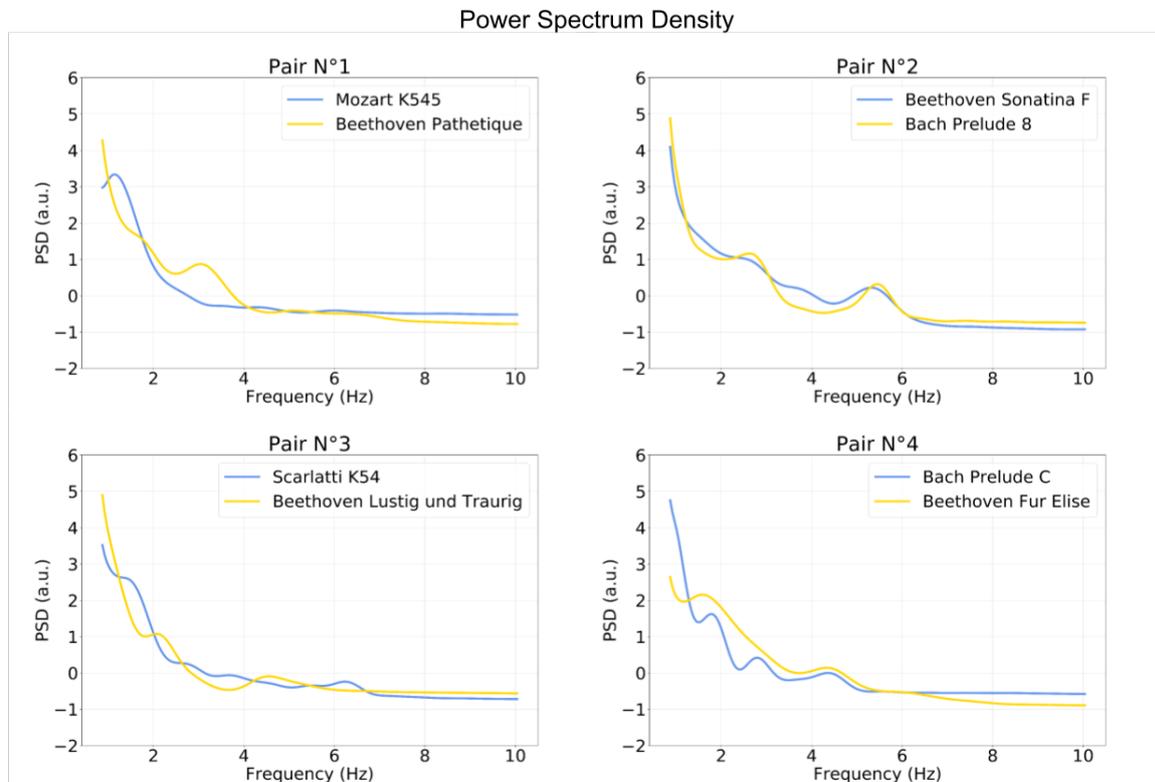


FIGURE 8: POWER SPECTRUM DENSITY OF THE ENVELOPE OF EACH MUSICAL EXCERPT. THE EXCERPTS WERE PRESENTED EITHER INDIVIDUALLY IN THE MONODIC CONDITION OR BY PAIR IN THE DICHOTIC CONDITION. PAIRS WERE CHOSEN IN ORDER TO MINIMIZE DIFFERENCES IN THE RHYTHMIC STRUCTURE.

PROCEDURE

Participants were comfortably seated in front of a 28" computer screen at a distance of approximately 80 cm. The task was divided in two parts: Monodic and Dichotic.

In the monodic part, participants were instructed to concentrate their attention on the musical excerpt as if they were trying to memorize it. Then, the musical excerpt was played for 60 seconds. During the stimulus presentation, a fixation cross was displayed at the screen center. Each of the musical excerpts was presented twice for a total of sixteen 60 second trials. The order of presentation avoided direct repetition of a given excerpt.

In the dichotic part, four pairs of musical excerpts were presented dichotically ($\pm 90^\circ$ from the azimuth). Participants were visually instructed which side (left or right) they had to attend. Each pair was presented four times (attend left, attend right for each of the two excerpts in the pair) for a total of sixteen 60 second trials. The whole experiment was monitored using a custom version of OpenSesame developed by Oticon Medical (Omexp: Patou et al., 2019).

COGNITIVE TASKS

In this study we were interested in three cognitive functions that are closely related to auditory scene analysis: WM, Attentional Inhibition and Sustained Attention. Attentional inhibition and sustained attention were measured online and from home for every participant using OpenSesame (Mathôt et al., 2012) and JATOS online test platform (Lange et al., 2015). This remote testing prevented an excessively long experimental session.

REVERSE DIGIT SPAN TASK

To measure WM, we used the Reverse Digit Span Task (RDST). The RDST is a classical WM task in which participants are presented with an auditory sequence of digits (between 1 and 9) that they have to immediately report orally in the reverse order. The length of the sequence increases until the participant makes two errors in a row. The test starts with a sequence of three digits and can reach a maximum of nine digits. The WM score corresponds to the size of the longest correctly reported sequence. The RDST was performed just before the listening task.

AUDITORY STROOP

We measured attentional inhibition with an online auditory Stroop task adapted from Donohue et al. (2012). The auditory stimuli consisted of French words “aigu” (“high”) and “grave” (“low”) spoken by a native French speaker, digitally sampled at 44100 Hz and resynthesized (Praat: Boersma, 2001) to obtain two tokens of each word with an average fundamental frequency of ~ 80 and 260Hz. All tokens lasted 300ms. Congruent trials consisted of the word “aigu” pronounced in a high pitch and the word “grave” pronounced in a low pitch while incongruent trials consisted of the word “aigu” pronounced in a low pitch and the word “grave” pronounced in a high pitch. Participants were instructed to decide as quickly and accurately as possible what was the pitch of the spoken word,

regardless of its meaning, by pressing on a keyboard. The response side to low or high pitch was counterbalanced across participants. The SOA was set to 1350 ms plus the time needed to respond.

A total of 260 congruent and 260 incongruent trials were presented in a pseudo-random order in 10 blocks of 52 trials (20 minutes).

For each participant we computed the accuracy (% of correct responses), the reaction times, the efficiency score for congruent and incongruent trials (accuracy/RTs) and an Inhibition score (Congruent efficiency - Incongruent efficiency). Thus, the higher the inhibition score, the better the attentional inhibition. The computation of these metrics allows one to take into account the Speed Accuracy Trade Off (SATO) (Gartner and Strobe, 2020).

AUDITORY SUSTAINED ATTENTION TO RESPONSE TASK (SART)

We measured sustained attention with an online auditory SART adapted from Seli et al. (2012).

The auditory stimuli consisted of digits (from 1 to 9) spoken by a native French speaker digitally sampled at 44100 Hz. Their duration was ~ 400 ms and they were randomly presented. Each digit was followed by a pink noise mask for 900ms. Participants were instructed to press the spacebar as quickly and accurately as possible each time they heard a digit except when they heard the digit 3. In this case, they had to withhold their answer. A total of 225 trials (200 Go and 25 NoGo) were undertaken by the participants, lasting around 7 minutes.

Accuracy for Go and NoGo trials were computed as well as RTs for Go trials. A sustained Attention score was computed as follows: $1000 * (\text{Accuracy_NoGo_Trials} / \text{Average_RT_Go_Trials})$. The higher the sustained attention score, the better the sustained attention. As for attentional inhibition, the computation of this complex metric allows one to take into account the SATO in SART performances (Seli et al., 2013, 2016; Hallion et al., 2020).

EEG DATA

Electroencephalography data were recorded using an ANT Refa8 amplifier and a 21-electrode cap (arranged in accordance with the International 10-20 system) with a

sampling rate of 256 Hz and an average reference. They were preprocessed using the MNE-python package (Gramfort et al., 2013) according to the procedure described in Crosse et al., 2021. First, raw EEG data were cut into 60s epochs, each epoch representing a trial, and the first 500ms of each epoch was discarded to avoid modeling the response to the stimulus onset. Next, EEG data were digitally filtered between 1-40Hz and bad channels were interpolated via spherical spline interpolation when necessary. Then, ICA was performed to remove eye-blinks and saccades artifacts. Afterward, EEG data were digitally filtered between 1-9Hz (Greenlaw et al., 2020; Biesman et al., 2015) using a 4th order Butterworth zero-phase-shift filter and downsampled to 64Hz. Finally, 60s epochs were cut into 30s epochs, resulting in 32 epochs for monodic and dichotic conditions, with the aim of having enough data to train and test the stimulus-reconstruction model.

AUDIO FEATURE EXTRACTION

Amplitude envelopes of the audio monodic stimuli were obtained using the function `human_cochleagram` from the Python `Pycochleogram` package (<https://github.com/mcdermottLab/pycochleogram>). This function allows 1) computing an Equivalent Rectangular Bandwidth (ERB) filter bank and 2) using this filter bank to decompose the signal into subband envelopes. Afterwards subband envelopes were averaged to obtain a unique envelope. Each envelope was then digitally filtered between 1-9Hz with a 4th order Butterworth zero-phase-shift filter, downsampled to 64Hz and cut into 30s long epochs in order to match EEG preprocessing.

STIMULUS RECONSTRUCTION

One objective of this study was to investigate if it was possible to reconstruct and classify monodic and dichotic natural polyphonic music from neural data. To do so, we used the same stimulus reconstruction approach as O'Sullivan et al. (2015) (this approach is also described in several other articles, see Crosse et al., 2016 and Crosse et al., 2021 for a comprehensive description).

GENERAL PROCEDURE

The stimulus reconstruction approach allows one to reconstruct an estimate of the envelope of the auditory stimulus s using electrophysiological neural data d via a linear reconstruction model g (O'Sullivan et al., 2015). The reconstruction model $g(\tau, n)$ is a temporal response function (see Lalor et al., 2009 for details) that maps neural data $d(t, n)$ to stimulus $s(t)$ as follow :

$$\hat{s}(t) = \sum_n \sum_{\tau} d(t + \tau, n)g(\tau, n)$$

Where \hat{s} denotes the reconstructed stimulus, $d(t, n)$ represent the response of electrode n at time $t=1 \dots T$ and τ are some time lags that represent a window in which the brain's response to the stimulus is supposedly optimal. We defined τ to go from 0.200ms pre-stimulus to 0.350ms post-stimulus based on previous works (O'Sullivan et al., 2015; Di Liberto et al., 2020b, 2021). At a sampling rate of 64 Hz, this corresponds to 36 sample shifts (incl. the zero shift).

The model g is estimated by minimizing the mean-squared error between the original stimulus $s(t)$ and the reconstructed one $\hat{s}(t)$:

$$\min e = \sum_t [s(t) - \hat{s}(t)]^2$$

A robust minimizer of the mean-squared error is obtained using the following matrix operations:

$$g = (D^T D + \lambda I)^{-1} D^T s$$

Where D is the lagged time series of the response matrix d and λ is a ridge parameter term introduced to avoid overfitting. The model parameters as well as the ridge parameter are generally estimated using leave-one-out cross validation (LOO) procedure. Once the model parameters have been tuned, it can be tested on new data (generally the leaved out fold) and the end of this process is the reconstructed stimulus \hat{s} . Finally, a reconstruction score is computed, to estimate the reconstruction accuracy of the model, as the Pearson correlation between the reconstructed stimulus \hat{s} and the original stimulus s . The higher the reconstruction score, the better the reconstruction accuracy.

TRAINING AND TESTING THE MODEL

In this study, we wanted to see if we can reconstruct a trial with a model trained on every other monodic/dichotic trial. To do so, we used a custom Python script where a first LOO was performed to separate training and test phase and because we introduced a ridge parameter (λ) we used a second LOO nested in the training phase to estimate the optimal value of the ridge parameter from a set of 20 logarithmically spaced values from 10^{-10} to 10^{10} .

At the end of the test phase, a reconstruction \hat{s}_i is obtained for each musical excerpt s_i , based on a model trained on all the other musical excerpts. This procedure allowed us to make sure that the musical excerpt we wanted to reconstruct was only used to validate the model.

Because we had 32 monodic and 32 dichotic trials, we trained separate models for each trial type. Therefore, this resulted in 32 models for monodic trials for each participant. For the dichotic trials, unlike O'Sullivan et al. (2015), we chose to train the models to reconstruct the attended musical excerpts only because 1) we were not interested in estimating which musical excerpt the participant was not attending to and 2) their results indicate that performances of unattended models are worse than performances of attended models. Thus, this resulted in 32 dichotic models per participant.

PERFORMANCES METRICS

RECONSTRUCTION ACCURACY

The reconstruction accuracy corresponds to the Pearson correlation between the reconstructed envelopes and the original musical envelopes. This led to 32 r values per participant. An average reconstruction accuracy score (r_{mono}) was also computed per participant.

Similarly, for dichotic trials, the reconstruction was computed for both attended (r_{Attended}) and unattended ($r_{\text{Unattended}}$) musical excerpts as the Pearson correlation between the reconstructed envelopes and the original attended and unattended musical envelopes.

Because we were interested in evaluating the increase in performance for attended vs the unattended stimuli, we defined a Dichotic Performance Index (DPI) as follows:

$$\text{DPI} = \frac{1}{N} \sum_{i=1}^N \frac{(r_{Ai} - r_{Ui})}{\sqrt{2}(r_{Ai}^2 + r_{Ui}^2)^{1/2}} \quad (\text{Def 1})$$

Where r_A and r_U are the r_{Attended} and $r_{\text{Unattended}}$ scores of each trial. The advantage of this measure is to provide for each participant a single measure summarizing the distance between attended and unattended r values while being robust to inter-trial variability.

CLASSIFICATION ACCURACY

Aside from reconstruction accuracy, we computed a classification accuracy score. For monodic trials, the classification accuracy score was the percentage of monodic trials that were correctly identified. We considered a trial to be correctly identified when r_{mono} was higher than all the correlations between the reconstructed musical excerpts envelope and the other musical excerpts envelope. Similarly, for dichotic trials, we considered a trial to be correctly identified when r_{Attended} was higher than $r_{\text{Unattended}}$. This score reflects the percentage of trials for which we were able to correctly classify the attended/unattended music, in dichotic trials.

STATISTICAL ANALYSIS

All statistical analyses were performed using the R software (R Core Team, 2022) with the `lme4` (Bates et al., 2015), `stats` (R Core Team, 2022), `car` (Fox, 2019) and `lmerTest` (Kuznetsova et al., 2017) packages.

NORMALITY TESTING, OUTLIERS REJECTION AND EFFECT SIZES

When appropriate, we assessed normality of the data distributions with a Shapiro-Wilk test. We tested for potential outliers using either a Grubbs test or a Tukey Fence for Outliers. We assessed heteroscedasticity using either an F test or a Fligner test. Finally, to measure the effect sizes we computed either a Cohen's d or a Rank-Biserial Correlation.

COGNITIVE FUNCTIONS

ONLINE TEST RELIABILITY

To test the reliability of online cognitive tasks, we compared Go/Congruent and NoGo/Incongruent distributions using a Wilcoxon rank sum test as differences between the two classes were strongly predicted by previous findings (REFs).

EFFECT OF MUSICAL EXPERTISE

To see whether musical expertise has an impact on cognitive functions, we compared the cognitive function test's performance of participants with low and high musical expertise using either a Student t test or a Wilcoxon rank sum test, as appropriate.

PERFORMANCE METRICS**INDIVIDUAL AND GROUP CHANCE LEVELS FOR CLASSIFICATION ACCURACY**

Individual, as well as group classification accuracy chance levels were estimated with an inverse cumulative probability distribution function for binomial distribution (O'Sullivan et al., 2015) and computed using the R `qbinom` function.

EFFECT OF MUSICAL EXPERTISE

To test whether musical expertise has an impact on reconstruction and classification accuracies we, again, compared the distributions of participants with low and high musical expertise using Student t test as they were all normal.

Additionally, to see if the distributions of participants with low and high musical expertise were different from the null distribution for reconstruction and classification accuracies, we used a One sample Student t test, with $\mu = 0$ and with $\mu = 50$, respectively.

COGNITIVE FUNCTIONS AND PERFORMANCES METRICS

In order to see if the chosen cognitive functions were good predictors of the reconstruction and classification accuracies, we used linear models computed on averaged data. A backward stepwise regression, where all the predictors (i.e., sustained attention, WM and attentional inhibition) were entered simultaneously, was used to select the best model for each dependent variable (i.e., reconstruction and classification accuracies) in each condition (monodic and dichotic). The best model was selected as the one with the lowest Akaike Information Criterion (AIC). After the best model was selected, models' assumptions of normality, homoscedasticity and independence of the residuals were confirmed using Jarque-Bera, Breush-Pagan and Durbin-Watson tests respectively. Finally, effect sizes were computed as the Cohen's f^2 .

RESULTS

The objectives of this study were: 1) to test if it was possible to reconstruct and classify monodic and dichotic natural polyphonic music from neural data, 2) to investigate the relation between specific executive functions and the reconstruction and classification

performances and 3) to explore the impact of musical expertise on these above-mentioned performances.

First, some results about cognitive functions are briefly exposed. Then, results about reconstruction and classification performances for both monodic and dichotic stimuli are presented. Next, results about the relation between executive functions and reconstruction and classification performances are exposed. Finally, the impact of musical expertise over the performances are reported.

COGNITIVE FUNCTIONS

Because we used an online version of the SART to measure sustained attention, we first compared the accuracy for Go and NoGo trials in order to see whether we could reproduce the findings reported in the literature. After outlier correction (see Methods), the test indicates a very large difference between the accuracy of Go ($M=99.3\%$ $sd=\pm 1.65\%$) and NoGo trials ($M=77\% \pm 14.8\%$) ($W = 1346.00$, $p < .001$; $r = 0.97$). This confirms the reliability of the online version of the SART.

We proceeded similarly with the data of the online auditory Stroop test, by comparing the global efficiency of congruent and incongruent trials. The test indicates a large positive difference between global efficiency of congruent ($M=166.7 \pm 28.5$) and incongruent trials ($M=151.9 \pm 24.4$) ($W = 1040.00$, $p = 0.021$; $r = 0.30$). This confirms the reliability of the online version of the auditory Stroop task.

Then, we explored whether musical expertise has an impact on the selected cognitive function, by comparing participants with low ($N=21$) and high ($N=19$) musical expertise. For WM, the effect of musical expertise was non-significant ($M\text{-low}=5.04 \pm 1.2$, $M\text{-high}=5.47 \pm 1.2$, $w=238$, $p = 0.287$, $r = 0.19$).

For sustained attention scores, the effect of musical expertise was also non-significant ($M\text{-low}=149.83 \pm 38.2$, $M\text{-high}=165.39 \pm 32.7$, $t(38)=1.37$, $p = 0.17$, Cohen's $d = 0.44$).

Finally, for attentional inhibition scores, the effect of musical expertise was again non-significant ($M\text{-low}=13.6 \pm 11.8$, $M\text{-high}=16.3 \pm 11.4$, $t(38)=0.73$, $p = 0.47$, Cohen's $d = 0.23$). Thus, overall, musical expertise did not have a significant effect on the cognitive variables we measured, although one can observe that values tended to be greater for participants with high compared to low musical expertise (see Figure 9).

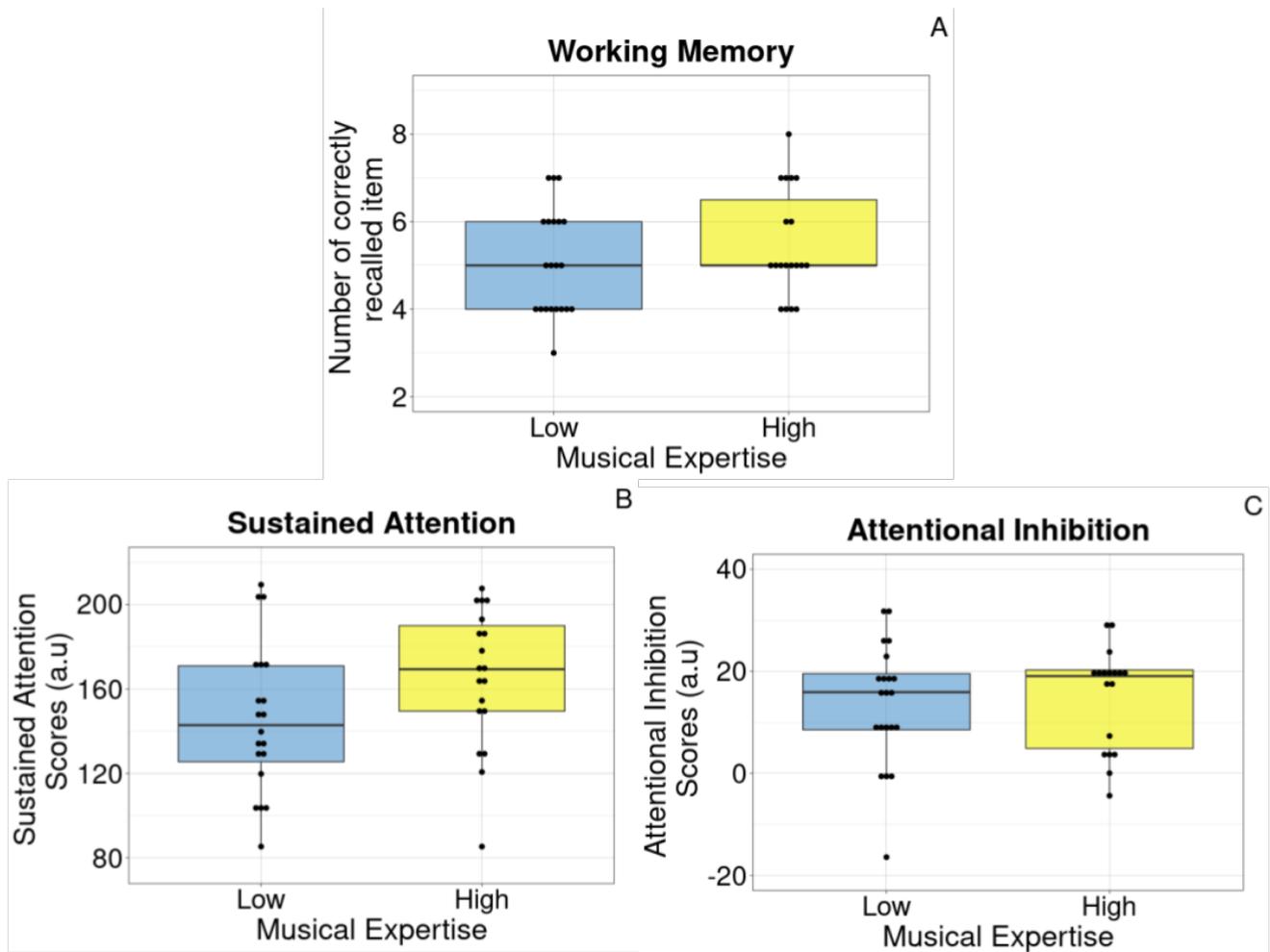


FIGURE 9: EFFECT OF MUSICAL EXPERTISE ON COGNITIVE FUNCTIONS. (A) WM SCORES FOR PARTICIPANTS WITH LOW MUSICAL EXPERTISE (LEFT) AND HIGH MUSICAL EXPERTISE (RIGHT). GREATER VALUES INDICATES BETTER WORKING MEMORY CAPACITY. (B) SUSTAINED ATTENTION SCORES FOR PARTICIPANTS WITH LOW MUSICAL EXPERTISE (LEFT) AND HIGH MUSICAL EXPERTISE (RIGHT). GREATER VALUES INDICATES BETTER SUSTAINED ATTENTION ABILITY. (C) ATTENTIONAL INHIBITION SCORES FOR PARTICIPANTS WITH LOW MUSICAL EXPERTISE (LEFT) AND HIGH MUSICAL EXPERTISE (RIGHT). GREATER VALUES INDICATE BETTER ATTENTIONAL INHIBITION. THE CENTRAL LINE OF THE BOX PLOTS REPRESENTS THE MEDIAN OF THE DISTRIBUTION, THE LOWER AND UPPER HINGES CORRESPOND TO THE FIRST AND THIRD QUARTILES AND THE WHISKERS REPRESENT THE LARGEST AND LOWEST VALUES NO FURTHER THAN 1.5 TIMES THE INTERQUARTILE RANGE FROM THE HINGES. INDIVIDUAL POINTS REPRESENT INDIVIDUAL DATA.

AAD PERFORMANCE

RECONSTRUCTION ACCURACY

One of the objectives of this study was to test if one can reconstruct monodic and dichotic natural polyphonic music from neural data, with a simple backward linear model.

On average, the reconstruction performance of the monodic musical excerpts was 0.063 (± 0.08). This result is significantly different from the null distribution ($t(40) = 14.32, p <$

0.001 ; Cohen's $d = 2.24$). Moreover, the average correlations between the reconstructions of the target excerpts (i.e., the listened excerpts) and their corresponding original envelopes were higher than all the correlations between the reconstructions of the target excerpts and the other original envelopes, as indicated by the higher values on the reconstruction matrix diagonal (figure 10 A).

For the dichotic stimuli, the average reconstruction performance was 0.028 (± 0.018) for the attended musical excerpts, 0.013 (± 0.013) for the unattended excerpts and 0.1 (± 0.16) for the DPI (Def 1). These reconstruction performances were clearly different from the null distribution: r_{Attended} ($w=856$, $p < 0.001$; $r = 0.99$), $r_{\text{Unattended}}$ ($w=749$, $p < 0.001$; $r = 0.83$). Importantly, attended streams were better reconstructed than unattended ones ($w=1191$, $p < 0.001$; $r = 0.45$). Moreover, for attended streams the average correlations between the reconstructions of the target excerpts (i.e., the attended excerpts) and their corresponding original envelopes were higher than all the correlations between the reconstructions of the target excerpts and the other original envelopes, as indicated by the higher values on the reconstruction matrix diagonal (see figure 10 B).

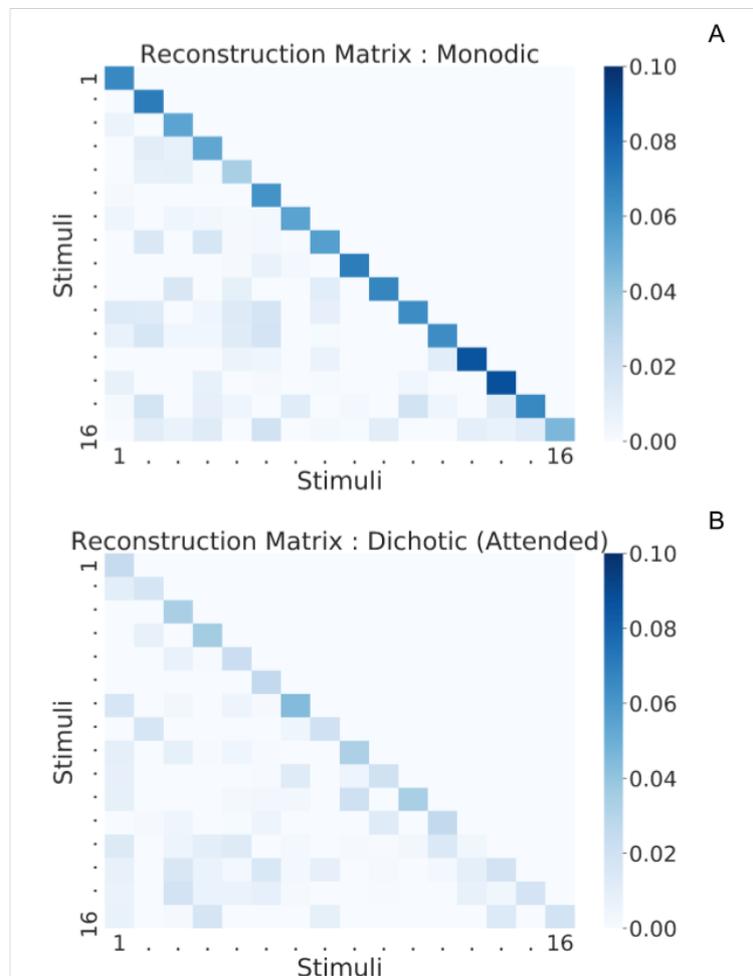


FIGURE 10: HEATMAPS OF RECONSTRUCTION SCORES FOR MONODIC (A) AND DICHOTIC (B) STIMULI. VALUES ON THE DIAGONAL INDICATE THE RECONSTRUCTION OF THE MUSICAL EXCERPT WITH THE CORRESPONDING ORIGINAL ENVELOPE (E.G. CORRELATION BETWEEN THE RECONSTRUCTION OF BEETHOVEN FUR ELISE AND THE ORIGINAL ENVELOPE OF BEETHOVEN FUR ELISE). IN (B) VALUES ON THE DIAGONAL INDICATE THE RECONSTRUCTION OF THE ATTENDED MUSICAL EXCERPT WITH THE CORRESPONDING ORIGINAL ENVELOPE. DARKER COLORS INDICATE HIGHER RECONSTRUCTION VALUES. RECONSTRUCTION SCORES WERE AVERAGED ACROSS REPETITIONS AND PARTICIPANTS. OVERALL, FOR BOTH MONODIC AND DICHOTIC ATTENDED STIMULI, RECONSTRUCTION VALUES ARE HIGHER WHEN COMPUTED WITH THE ORIGINAL ENVELOPE RATHER THAN WITH THE ENVELOPE OF ANOTHER STIMULUS.

Figure 11 shows a great inter-individual variability in the reconstructions of both monodic and dichotic musical excerpts.

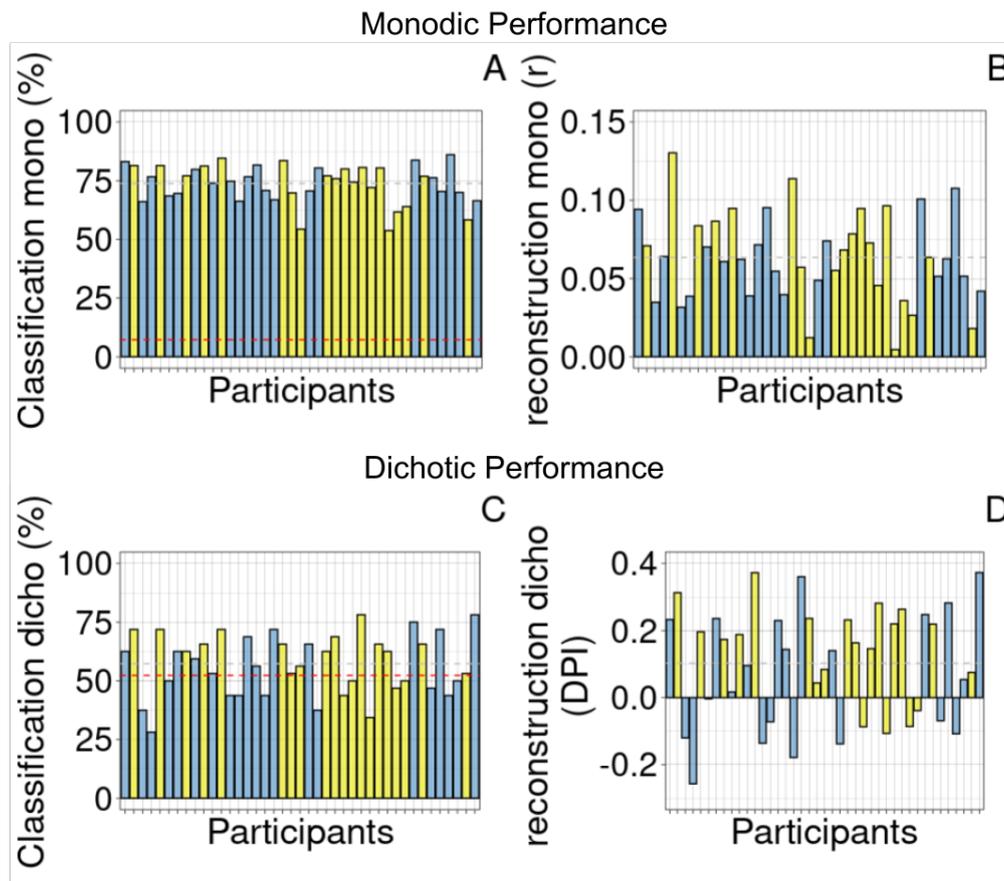


FIGURE 11: ILLUSTRATION OF MONODIC AND DICHOTIC RECONSTRUCTION AND CLASSIFICATION ACCURACY. (A) CLASSIFICATION PERFORMANCE FOR MONODIC STIMULI (AVERAGE = 73.8%). (B) RECONSTRUCTION PERFORMANCES OF MONODIC STIMULI (AVERAGE = 0.063). (C) CLASSIFICATION PERFORMANCE FOR DICHOTIC STIMULI (AVERAGE = 57.31%, GROUP CHANCE LEVEL = 52.3%). (D) DICHOTIC PERFORMANCE INDEX (DPI) OF DICHOTIC STIMULI (AVERAGE = 0.102), REPRESENTING THE DISTANCE BETWEEN ATTENDED AND UNATTENDED R VALUES. GRAY DASHED LINES INDICATE AVERAGE RECONSTRUCTION AND CLASSIFICATION. RED DASHED LINES INDICATE THE GROUP CHANCE LEVEL. BLUE-FILLED BARS REPRESENT PARTICIPANTS WITH LOW MUSICAL EXPERTISE WHILE YELLOW-FILLED BARS REPRESENT PARTICIPANTS WITH HIGH MUSICAL EXPERTISE.

We were also interested to see whether the monodic reconstruction performance could explain the dichotic one (Figure 12). Therefore, we computed a Pearson correlation that indicates a non-significant relation between the two performances ($r = 0.2$, $p\text{-value} = 0.197$).

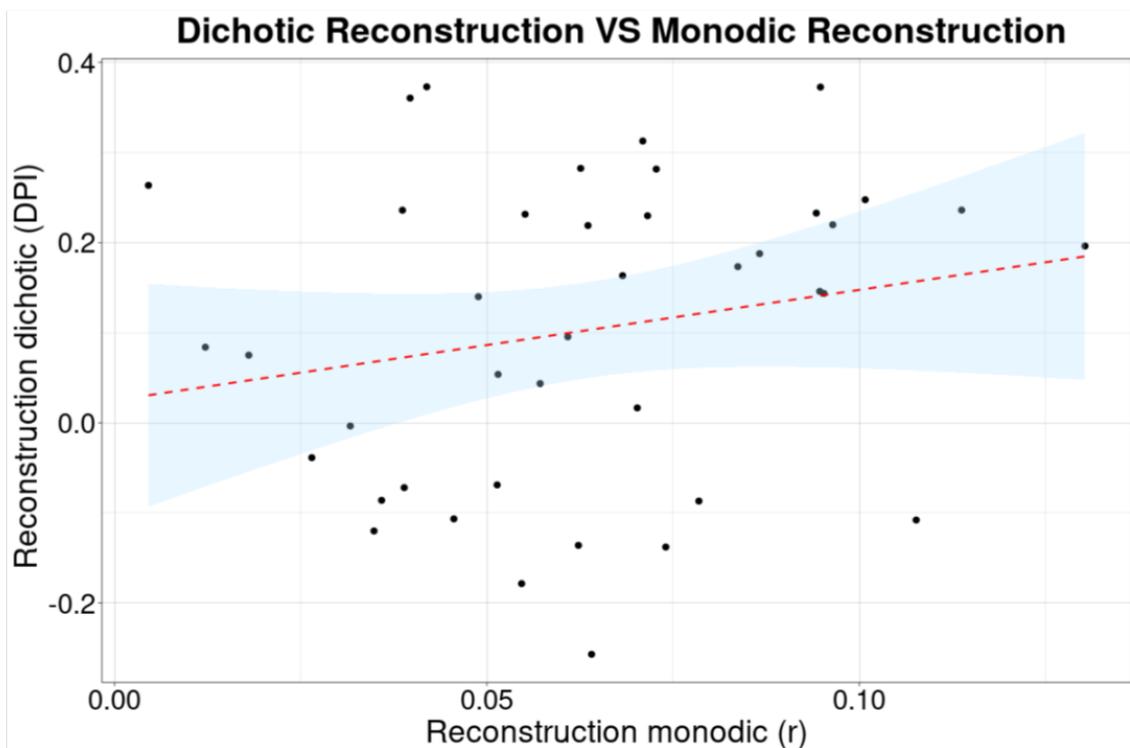


FIGURE 12: SCATTERPLOT ILLUSTRATING THE LACK OF SIGNIFICANT RELATION BETWEEN DICHOTIC AND MONODIC RECONSTRUCTION. ADJUSTED- $R^2 = 0.017$, P-VALUE = 0.2. RED DASHED LINE INDICATES THE REGRESSION LINE. BLUE AREA INDICATES CONFIDENCE INTERVAL AT 95%.

CLASSIFICATION ACCURACY

In addition to the reconstruction accuracy, we also investigated to what extent we could correctly classify the stimuli (Figure 11).

On average, classification accuracy for monodic trials reached 73.8% (± 8.22) and all the participants had a classification accuracy significantly greater than chance level (=12.5%, estimated using an inverse binomial test; see material and method). At the group level the average classification accuracy was clearly higher than the group chance level (=7.3%, estimated using an inverse binomial test).

As expected, the classification accuracy of dichotic trials was lower and reached 57.31% (± 12.8). At the individual level, only 10 participants out of 41 had a classification accuracy significantly greater than chance level (=65.65%, estimated using an inverse binomial test). However, at the group level, the average classification accuracy was higher than the group chance level (=52.3%, estimated using an inverse binomial test).

EFFECT OF MUSICAL EXPERTISE ON RECONSTRUCTION AND CLASSIFICATION

ACCURACIES

We were also interested in knowing whether the degree of musical expertise could have an impact on both reconstruction and classification accuracy of monodic and dichotic stimuli (Figure 13).

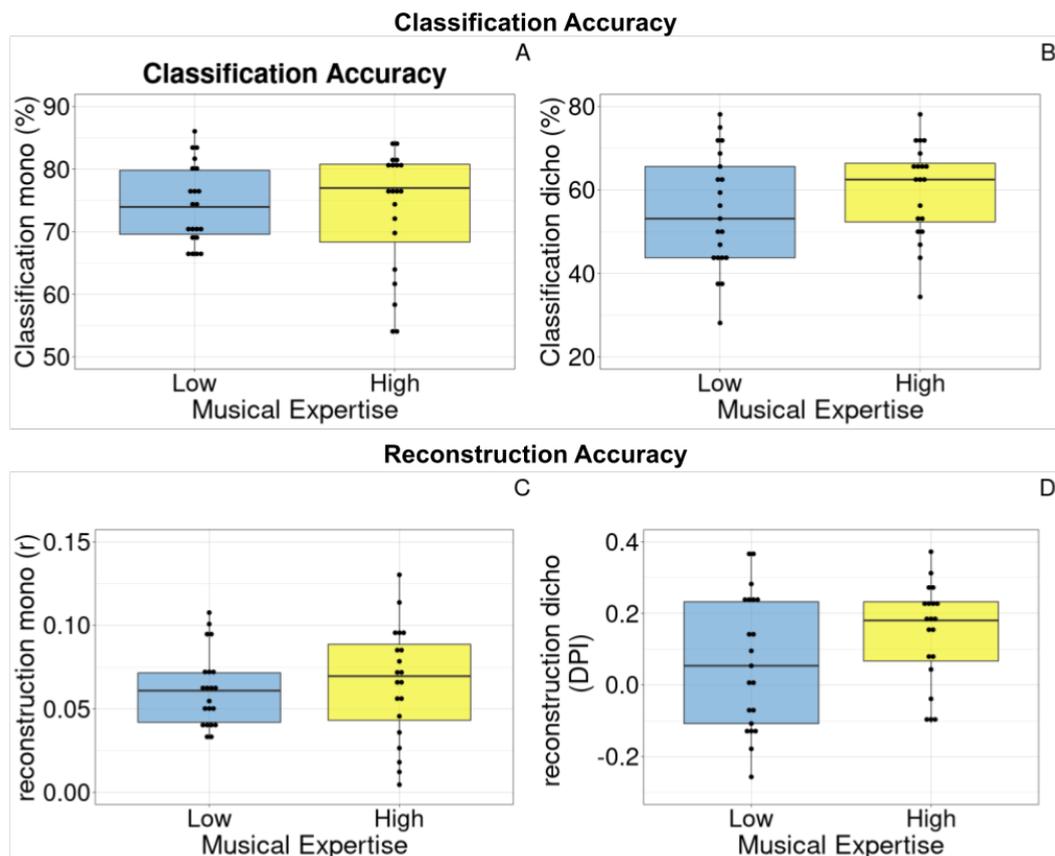


FIGURE 13: EFFECT OF MUSICAL EXPERTISE ON MONODIC AND DICHOTIC RECONSTRUCTION AND CLASSIFICATION. (A) CLASSIFICATION OF MONODIC STIMULI FOR LOW (LEFT) AND HIGH MUSICAL EXPERTISE (RIGHT). (B) CLASSIFICATION OF DICHOTIC STIMULI FOR LOW (LEFT) AND HIGH MUSICAL EXPERTISE (RIGHT). (C) RECONSTRUCTION OF MONODIC STIMULI FOR LOW (LEFT) AND HIGH MUSICAL EXPERTISE (RIGHT). (D) DPI OF DICHOTIC STIMULI FOR LOW (LEFT) AND HIGH MUSICAL EXPERTISE (RIGHT). THE CENTRAL LINE OF THE BOX PLOTS REPRESENTS THE MEDIAN OF THE DISTRIBUTION, THE LOWER AND UPPER HINGES CORRESPOND TO THE FIRST AND THIRD QUANTILES AND THE WHISKERS REPRESENT THE LARGEST AND LOWEST VALUES NO FURTHER THAN 1.5 TIMES THE INTERQUARTILE RANGE FROM THE HINGES. INDIVIDUAL POINTS REPRESENT INDIVIDUAL DATA.

We found a clear difference between the null distribution ($\mu = 50$) and both the DPI and dichotic classification accuracy of high musical expertise participants: DPI ($t(19)=4.64$, $p < 0.001$; Cohen's $d = 1.04$), Dichotic classification accuracy ($t(19)=4$, $p < 0.001$; Cohen's $d = .89$). However, we found no difference between the null distribution and the dichotic

performances of low musical expertise participants: DPI ($t(20)=1.54$, $p = 0.14$, Cohen's $d = 0.34$), Dichotic classification accuracy ($t(20)=1.56$, $p = 0.13$; Cohen's $d = 0.34$).

However, the direct effect of musical expertise was significant neither for monodic (M-low= 0.061 ± 0.022 , M-high= 0.065 ± 0.034 , $t(39)=0.42$, $p = 0.67$; Cohen's $d = 0.13$) nor for dichotic reconstruction (M-low= 0.063 ± 0.18 , M-high= 0.144 ± 0.13 , $t(39)=1.56$, $p > 0.12$; Cohen's $d = 0.49$) and not either for both monodic (M-low= 74.22 ± 6.34 , M-high= 73.42 ± 9.8 , $t(39)=0.31$, $p = 0.75$; Cohen's $d = 0.1$) nor for dichotic classification (M-low= 54.8 ± 13.9 , M-high= 60 ± 11.2 , $t(39)=1.32$, $p = 0.19$; Cohen's $d = 0.41$).

EFFECT OF COGNITIVE FUNCTIONS ON RECONSTRUCTION AND CLASSIFICATION

ACCURACIES

The main goal of the present study was to test whether selected cognitive functions could explain a part of the inter-individual variability of the reconstruction and classification performances during dichotic listening as well as during the listening of monodic stimuli (Figure 14).

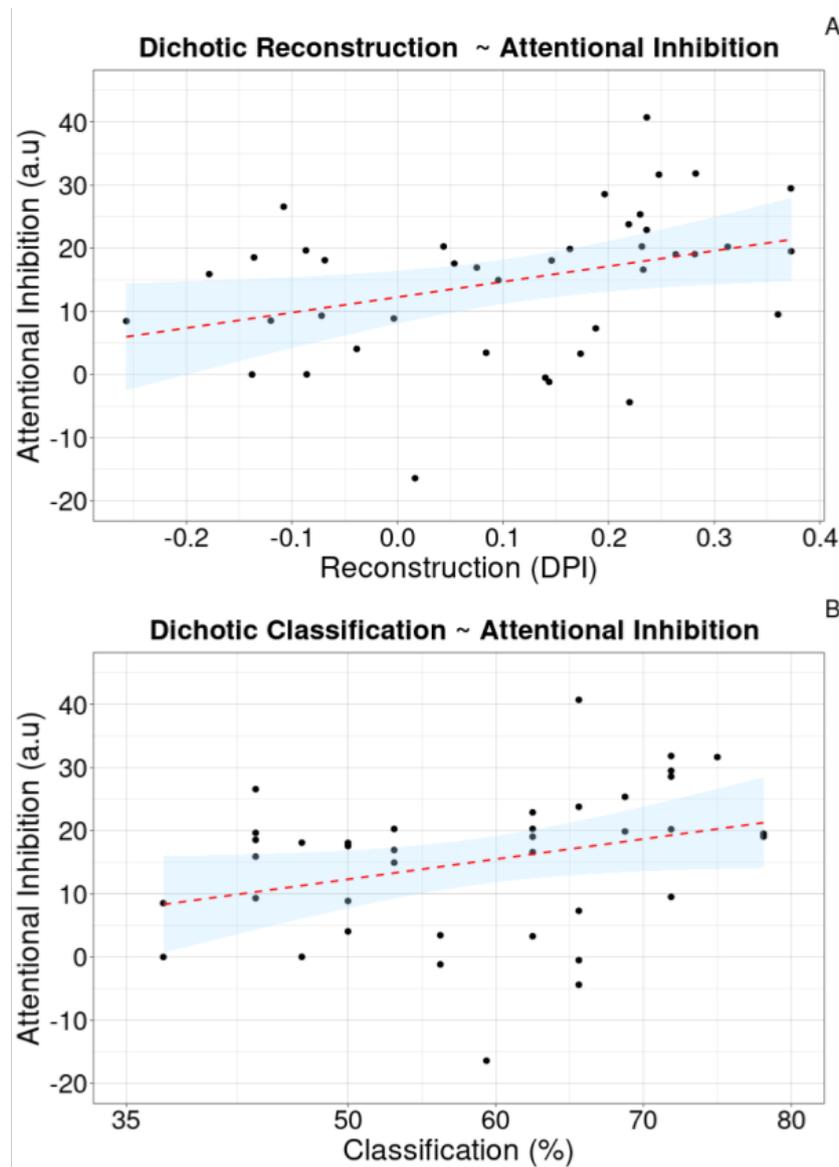


FIGURE 14: ILLUSTRATION OF THE SIGNIFICANT RELATION BETWEEN ATTENTIONAL INHIBITION AND DICHOTIC LISTENING METRICS. (A) LINEAR REGRESSION OF THE DPI BY ATTENTIONAL INHIBITION. ADJUSTED- $R^2 = 0.1$, P-VALUE = 0.023. (B) LINEAR REGRESSION OF THE DICHOTIC CLASSIFICATION ACCURACY BY ATTENTIONAL INHIBITION. ADJUSTED- $R^2 = 0.08$, P-VALUE = 0.039. RED DASHED LINE INDICATES THE REGRESSION LINE. BLUE AREA INDICATES THE CONFIDENCE INTERVAL AT 95%. GREATER VALUES INDICATE BETTER ATTENTIONAL INHIBITION.

ATTENTIONAL INHIBITION

For monodic stimuli we found no effect of attentional inhibition on reconstruction accuracy ($t(38)=1.88$, $p > 0.068$) and the prediction model was not significant (adjusted- $R^2 = .06$, $f(1,38)=3.52$, $p = 0.068$, Cohen's $f^2 = 0.09$). There was also no effect of attentional inhibition on classification accuracy ($t(38)=1.42$, $p = 0.16$). Once again the prediction model was not significant (adjusted- $R^2 = .03$, $f(1,38)=2.01$, $p = 0.16$, Cohen's $f^2 = 0.05$).

For dichotic stimuli, we found a medium positive effect of attentional inhibition on DPI ($t(38)=2.36$, $p = 0.024$) and attentional inhibition predicted around 10% of the DPI (adjusted- $R^2=.1046$, $f(1, 38)=5.55$, $p = 0.024$, Cohen's $f^2 = 0.15$). This regression model was significantly different from the null model ($F(2,39) = 5.55$, $p = 0.023$). We also found a small but significant positive relation between attentional inhibition and classification accuracy ($t(38)=2.13$, $p = 0.039$). Attentional inhibition predicted around 8% of the classification accuracy (adjusted- $R^2 = .0835$, $f(1,38)=4.56$, $p = 0.039$, Cohen's $f^2 = 0.12$). This regression model was significantly different from the null model ($F(2,39) = 4.55$, $p = 0.039$).

Additionally, no significant interaction was found between attentional inhibition and any of the other cognitive variables.

OTHER COGNITIVE FUNCTIONS

For WM and sustained attention we found no effect on monodic nor dichotic reconstruction ($t(38) < 2$, $p > 0.05$) nor classification accuracy either ($t(38) < 2$, $p > 0.05$).

Using a backward stepwise regression (see Material and Methods), we compared a full model with all the cognitive functions as predictors, to a simple model with attentional inhibition only for both dichotic reconstruction and classification accuracy. The AIC criteria of the simple model were lower (DPI = -145.1 and classification = 200) than the one of the full model (DPI = -142.2 and classification = 202.7) suggesting that the full model was not significantly better than the simple one.

DISCUSSION

In the present study we used a reconstruction stimulus approach to explore whether it was possible to reconstruct and classify both monodic and dichotic natural polyphonic musical excerpts on the basis of EEG recordings. We also measured sustained attention, attentional inhibition and working memory via cognitive tests either online (SART and Stroop tests) or with a more traditional approach (Reverse Digit Span Task). Then, we employed linear regression models to see whether cognitive functions and musical expertise explain reconstruction and classification performance of neural data. Additionally, we explored whether musical expertise affects executive functions as well as neural tracking of musical stimuli by comparing non-musical and musical experts.

First, our results show that it is possible to reconstruct both monodic and dichotic natural polyphonic excerpts based on neural data using the reconstruction stimulus approach (see Figure 3). In particular, for the dichotic stimuli, attended stimuli were better decoded than unattended stimuli. Moreover, it was also possible to classify not only the monodic stimuli with good precision (73.8%), but also the dichotic stimuli although to a lesser extent (57.3%).

Second, unlike what we expected, we found no difference in terms of cognitive functions between participants with high and low musical expertise. Moreover, we found no effect of musical expertise on reconstruction and classification performances for monodic nor dichotic stimuli.

Finally, and most importantly, while for monodic stimuli we found no effect of the selected cognitive functions over reconstruction and classification accuracies, for dichotic stimuli attentional inhibition did explain around 10% of the reconstruction accuracy and around 8% of the classification accuracy. No other cognitive function was a significant predictor neither of dichotic reconstruction nor of classification accuracies.

We will first discuss the feasibility using ecological musical stimuli within the auditory attention decoding approach. Then, we will discuss the relation between cognition and performances of stimulus reconstruction. Finally, we will consider the effect of musical expertise over the chosen algorithm performances.

In this work, we explored whether it was possible to both reconstruct and classify monodic and dichotic ecological musical excerpts based on neural data.

In the monodic condition, reconstruction performances were not very high (0.064 on average) but significantly greater than a null distribution. The reconstruction matrix and the resulting diagonal (Figure 2), indicate that the average correlation between the reconstructed musical excerpts and their corresponding original envelopes were clearly higher than the average correlations between the reconstructed musical excerpts and all the other envelopes. This result indicates that, at the group level, no matter their magnitudes, the reconstructions are more similar to the original envelopes than to the other ones. In addition, reconstruction values using polyphonic music stimuli were close to the results obtained in other similar studies using monodic music stimuli (e.g. Di Liberto et al., 2020b), although somewhat smaller than when using multiway canonical

correlation analysis (Di Liberto et al. 2020a, 2021). In the dichotic condition, as expected, reconstruction performances were lower (0.028 on average for attended stimuli, 0.013 for unattended ones and 0.1 on average for the DPI). However, reconstruction was significantly greater than chance for both attended and unattended stimuli. Moreover, the greater reconstruction performances for attended compared to unattended stimuli shows that participants were able to do the task, namely concentrate their attention on the target stream and inhibit the concurrent one, despite the reported difficulty to do so. Overall, these results confirm previous findings showing that the brain tracks the amplitude envelope of target and concurrent streams (e.g., Ding and Simon, 2012; O'Sullivan et al., 2015). Importantly, the replication of these previous findings is a crucial test to confirm that the design and analysis pipeline were appropriate to reconstruct both monodic and dichotic realistic polyphonic musical stimuli based on neural data.

Classification performance can be interpreted in a similar line. In the monodic condition, classification of the attended musical excerpt was good (around 74%) and consistent with previous findings (Schaefer et al., 2011). In addition, all participants showed a classification accuracy clearly higher than chance level (7.3% for a thirty-class problem). For the dichotic condition, the accuracy to classify the target stream was lower (57%), but still higher than chance at the group level (= 52%). This is consistent with the work of Cantisani and colleagues (2019) who found a classification accuracy of around 58% for duet stimuli. However, while their duet stimuli were composed of two different monophonic instruments, we used only pairs of polyphonic piano excerpts. This leads to a more complicated auditory scene in our experiment because 1) in a single excerpt, the piano plays several notes at a time, 2) the timbre of two different pianos is rather similar. Overall, the results discussed here indicate that, using a reconstruction stimulus strategy, it was possible not only to detect which musical excerpt between many was presented in a monodic condition but also to detect which musical excerpt was attended in a rather challenging dichotic listening context.

However, it is important to consider that interindividual variances of the reconstruction and classification performances in both conditions were relatively high, in particular in the dichotic condition. Indeed, the DPI ranged from ~ -0.28 to ~ 0.39 and the classification from almost 30% to around 80% in dichotic listening. One possible explanation of such variability could be the different quality of EEG data across participants. Nevertheless,

because classification and DPI are not sensitive to the magnitude of the reconstruction values, this variability should probably not be attributed to the data quality. Indeed, classification accuracy is computed as the percentage of trials for which r_{Attended} is higher than $r_{\text{Unattended}}$. Moreover, the absolute values of r_{Attended} and $r_{\text{Unattended}}$ do not affect the computation of the DPI (see Material and Methods section). Therefore, these metrics should not be biased by inter-trial and inter-participant variability caused by data quality fluctuations. This hypothesis is also corroborated by the non-significant correlations between monodic reconstruction and DPI ($r=0.2$, $p\text{-value}=0.197$, see Figure 5) and between the reconstruction of monodic and both Attended ($r=0.22$, $p\text{-value}=0.16$) and Unattended dichotic stimuli ($r=0.06$, $p\text{-value}=0.66$). Indeed, if one considers the monodic condition to be the baseline, then the absence of correlation between the baseline and the experimental condition (i.e., dichotic listening) suggests that factors such as skull thickness, brain orientation or Signal to Noise Ratio are not likely to explain the dichotic performances. Importantly, it also provides hints on the fact that dichotic AAD performance seems to be dependent on other factors than just the general ability to listen to a musical piece.

One possible factor that we assessed in the present work is musical expertise. When we compared the average dichotic reconstructions and classifications distributions of low and high musical expertise participants against the null distribution, we observed a difference for participants with high musical expertise but not for participants with low musical expertise. However, we found no main effect of musical expertise on both dichotic and monodic performances. Put together, these findings suggest that in the most challenging condition, i.e., the dichotic condition, musical expertise may have a facilitatory effect, but the effect may be small. This result partially corroborates previous works showing a better cortical tracking of music stimuli for musicians compared to non-musicians (Di Liberto et al., 2020b; Doelling and Poeppel, 2015; Puschmann et al., 2019; Harding et al., 2019) nor the one showing better performances of musicians in dichotic listening (e.g., Barbosa-luis et al., 2021).

A possible reason for the small impact of musical expertise may rely on the importance of inhibiting irrelevant streams in our task. Indeed, a very recent study demonstrated that musical training was not associated with attention skills in a task where participants had to encode a target melody while inhibiting a distracting one (Blain et al., 2022, see also

Bidelman and Yoo, 2020 for a similar conclusion with speech-in-noise task). While these results are in line with our findings, they should be nonetheless shadowed by the fact that this type of task also requires processes that are supposedly better in musicians such as stream segregation (e.g., François et al., 2014), cognitive control (e.g., Bialystok et al., 2009) or WM (e.g., Zuk et al., 2014). That being said, the superiority of musically trained people in executive functions is still under debate (e.g., Boebinger et al., 2015; Schellenberg, 2011 and Sachs et al., 2017 for evidence in children) and we found no clear differences between low and high musical experts in cognitive scores here. Another possible reason may rely on the choice of the stimuli. In the current setup participants heard simultaneously a different musical excerpt in each ear. However, the two excerpts were rather controlled in terms of rhythmic structure and spectral features (see Figure 1). Thus, while participants could rely on the spatial dimension to segregate the two streams, the spectral and temporal differences between the musical excerpts presented dichotically were small. By limiting the use of spectral and temporal cues in stream segregation, this may have reduced musicians' advantage (Parbery-Clark et al., 2012; Fuller et al., 2014; Parbery-Clark et al., 2009a, 2009b).

The main objective of this work was to explore whether executive functions, and more specifically WM, sustained attention and attentional inhibition, could explain the variability of reconstruction and classification of ecological monodic and dichotic stimuli based on neural data. We wanted to see whether the ability to sustain attention would impact the auditory tracking of the stimulus to attend. More precisely, participants with weak sustained attention abilities may experience more attentional disengagement periods (toward the stimulus to be attended), leading to poorer auditory cortical representation of the stimulus and therefore poorer AAD performance. However, our analysis revealed that sustained attention was not related to the performances in monodic and dichotic conditions. At least two non-exclusive explanations may be discussed here: the choice of the attentional task and the duration of the musical stimuli. First, as suggested by Helton (2009), SART performances, and more specifically errors of commission, can be seen as failures of motor control, while motor control may not be crucial when sustaining attention of auditory objects. Relatedly, compared to the continuous stream of music, the discrete stream of the SART may lead to differences in attentional maintenance and attentional capture (Yantis and Jonides, 1984; Sturm and

Willmes, 2001; Esterman et al., 2013; Esterman and Rothlein, 2019). Second, the common scores derived from the SART do not reflect moment-to-moment lapses of attention because they are mainly global measures of sustained attention over long periods of time. Thus, it could be that the SART was not able to capture the smaller scale sustained attention variations that were relevant when listening to the one-minute musical excerpts we used.

The WM task yielded similar results with both the monodic and the dichotic listening conditions. This result is less surprising. Indeed, it seems possible that in our listening paradigm where stimuli were long ecological polyphonic music excerpts, the ability to follow the stream of interest strongly relied upon attentional inhibition and to a lesser extent upon rehearsal, encoding or updating mechanisms. Although this seems evident for dichotic stimuli, one should recall that even in the monodic condition, the use of polyphony implies auditory stream segregation and integration (Wright and Bregman, 1987). Importantly, the high ambiguity of polyphonic music (Pressnitzer et al., 2011) is possibly mediated by an interaction of both bottom-up perceptual and top-down attentional processes (Bigand et al., 2000; Disbergen et al., 2018). While it has been proposed that a crucial function underlying the stable maintenance of items in WM is the ability to deal with distractors (Conway et al., 2001; Cowan et al., 2005; Bledowski et al., 2010), the WM task that we used (i.e., Backward Digit Span Task) does not measure this inhibitory component. This may explain why the WM score failed to predict the decoding performance in a task requiring complex auditory scene analysis. Clearly, this does not mean that WM is not involved in both monodic and dichotic listening (Peretz and Zatorre, 2005; Escobar et al., 2020; Colflesh and Conway, 2007; James et al., 2014; Janata et al., 2002). Further work is needed to more precisely understand how the different components of WM are involved in such complex situations such as listening to ecological polyphonic musical stimuli.

This interpretation seems to be comforted by the fact that attentional inhibition explained around 10% of the variance of dichotic reconstruction performance and around 8% of the classification one. This result confirms our hypothesis that a better ability to resist external distractors leads to better neural tracking of the auditory targets. This extends previous findings showing that cognitive control is involved in dichotic listening tasks (Hugdhal et al., 2009; Falkenberg et al., 2011) and auditory scene analysis

(Goll et al., 2012). Moreover, it seems to corroborate the manifold proposal about the implication of an inhibitory process in dichotic listening (see for example Conway et al., 2001 or Bledowski et al., 2010). The specificity of the link between auditory attentional inhibition and dichotic listening is also supported by the non-significant correlation between attentional inhibition score and performance during monodic listening, that is with no distractors. Overall, this result suggests that an attentional inhibitory mechanism is at work during dichotic listening and possibly affects the way the brain tracks the relevant auditory information. This extends to ecological music stimuli previous findings of auditory selective attention studies showing a decreased brain response of non-relevant simple sounds compared to relevant ones (Chait et al., 2010; Alain and Woods, 1994; Berman et al., 1989; Bidet-Caulet et al., 2007, 2010; Michie et al., 1990, 1993). In addition, this result is also in line with the aforementioned findings suggesting that maintaining a stable representation of the target stream during dichotic listening could be mainly based on the ability to deal with distractors. Nevertheless, further work has to be done to exactly characterize what type of inhibitory process is the most important during dichotic listening. Indeed, with the Stroop task we were only able to measure distractor resistance but it is not excluded that intrusion resistance could play a role in dichotic listening as well. Importantly, this is the first evidence that attentional inhibition, as assessed by a cognitive task, may have a predictive power of the performance of a neuro-steered hearing aid. This result also raises the question of whether training inhibition abilities may improve auditory neural tracking in complex auditory scenes.

Overall, the results of the presented study showed the feasibility of using the reconstruction stimulus approach with ecological polyphonic musical stimuli. Importantly, they point to specific cognitive mechanisms being at work during dichotic musical listening in normal hearing listeners. Further research should be carried on to investigate other cognitive functions that are involved in streaming (e.g., divided attention) as well as more precise sub-components of the executive functions, like distractor and intrusion resistance or updating and rehearsal in WM, in order to see to what extent, they can predict the performance of the reconstruction stimulus approach. Also, it would be of interest to explore whether more general cognitive abilities such as general intelligence or cognitive flexibility could also be good predictors of AAD performance. In addition, it may be valuable to replicate this work with other populations

such as CI implanted, children or the elderly, to see if the findings are consistent. Finally, as we show the predictability power of specific executive functions, it appears conceivable to study the impact of executive-function-based therapies to improve neuro-steered hearing aids performance.

In conclusion, one can say that cognitive factors seem to impact AAD performance and taking advantage of this relation could be useful to improve next-generation hearing aids. Thus, the combination of specific cognitive training and neuro-steered hearing aids seems to be a serious option to help deaf people perceive the world a little more naturally.

STUDY TWO: THE EFFECT OF FAMILIARITY ON NEURAL TRACKING IS MODULATED BY THE MIND WANDERING

This chapter is taken from the article “The effect of familiarity on neural tracking is modulated by the mind wandering” from Belo, Clerc and Schön, submitted in 2022 in Behavioral Brain Research. This study is based on the same data as the Study One.

INTRODUCTION

It is well established now that sounds are represented in the brain at multiple levels including subcortical (Kraus and Chandrasekaran, 2010) and cortical areas (Nourski et al., 2009; Giraud and Poeppel, 2012). Notably, it has been shown that, at the cortical level, the brain tracks the amplitude envelope (AE) of auditory stimuli (Mesgarani et al., 2009; Kubanek et al., 2013). One way to investigate the cortical tracking of AE of continuous auditory stimuli is to use the stimulus reconstruction approach (SR) which allows reconstructing the AE of an auditory source using neural activity recorded with magneto/electroencephalography (Mesgarani et al., 2009). Moreover, SR is interesting to gain insights about the factors that may influence the cortical representation of continuous auditory stimuli, such as speech or music. For instance, studies have investigated the effect of speech intelligibility on the cortical tracking of speech stimuli (e.g., Baltzell et al., 2017; Peelle et al. 2013), the impact of musical expertise on speech and music representation (e.g., Di Liberto et al., 2020b; Doelling and Poppel, 2015; Hardin et al., 2019; Belo et al., 2022), or even the effect of attention in more complex auditory environments on the cortical representation of attended and unattended speech streams (e.g., Mesgarani and Chang, 2012; Ding and Simon, 2012; O’Sullivan et al., 2015; Zion Golumbic et al., 2013). However, it is possible that other factors, cognitive as well as behavioral, may impact the cortical representation of sounds (Belo et al., 2022; Ciccarelli et al., 2019).

One possible candidate is the familiarity with the stimulus. This idea is supported by studies showing that familiar words (Connine et al., 1990) and voices (Skuk and Schweinberger, 2013) are recognized more promptly than unfamiliar ones. This is further supported by other works showing distinct processing for familiar and unfamiliar voices (e.g., van Lancker and Kreiman, 1987; Blank et al., 2014). Very recent studies directly investigated the effect of familiarity on the cortical representation of speech and

found better cortical representation when listeners are more familiar with the stimuli (Cervantes et al., 2018; Baltzell et al., 2017). While these results are rather consistent for speech stimuli, this is not the case when investigating the role of familiarity on the cortical representation of musical stimuli. Indeed, some studies show, as described with speech stimuli, an enhancement of cortical tracking for familiar compared to unfamiliar song utterances (Vanden Bosch der Nederlanden et al., 2022) and music (Madsen et al., 2019, Weineck et al., 2022). However, other studies show the opposite pattern, meaning a better cortical tracking of unfamiliar music compared to familiar ones (Kumagai et al., 2017, 2018; Meltzer et al., 2015).

Beyond stimulus familiarity, another potential factor impacting cortical representations of auditory sources is the resistance to internal distractors. One way to address one's ability to resist internal distractors is to estimate the level of mind wandering during a task. While there is no clear definition of mind wandering (see Seli et al., 2018 and Christoff et al., 2018), one can consider it as periods of inattention (toward the task at hand) during which self-generated thoughts arise (Smallwood and Schooler, 2015). Interestingly, it has been shown that when the mind wanders, attention is reduced toward external stimuli, which negatively impacts the processing of external informations (Barron et al., 2011; Kam et al., 2011, 2014) and the achievement of the external task (Kam and Handy, 2013; Smallwood, 2013). Considering that cortical tracking is enhanced when attending the stimuli (e.g., O'Sullivan et al., 2015), periods of inattention toward the auditory stimuli (i.e., mind wandering) may negatively impact its cortical representation.

In light of these considerations, the objective of this study was to investigate to what extent familiarity and mind wandering may predict the reconstruction accuracy of musical ecological stimuli.

MATERIAL AND METHODS

PARTICIPANTS

A total of forty-one subjects took part in the experiment (28.93 ± 10.94 years, min = 18 years, max = 54 years, 22 females). All the participants included in the analysis reported no history of hearing disorders, attention impairment or neurological disorder and were not under medication.

The experiment was approved by the Operational Committee for the Evaluation of Legal and Ethical Risks (COERLE) of INRIA and was undertaken in accordance with the Declaration of Helsinki. Each participant provided written informed consent and was financially compensated.

MAIN EXPERIMENTAL TASK

STIMULI

Monodic stimuli consisted of 60-second recordings of four well-known and four less-known classical piano pieces. The well-known musical excerpts were taken from Beethoven's Für Elise, Bach's Prelude in C Major, Mozart's Sonata in C Major K545 and Beethoven's Pathétique Sonata. The less-known musical excerpts were taken from Beethoven's Lustig und Traurig, Scarlatti Sonata in A Minor K54, Beethoven's Sonata in F Presto and Bach's Prelude no°8. All musical excerpts were RMS normalized. Stimuli were presented via Sennheiser HD-25 supra-aural headphones at a comfortable level.

PROCEDURE

Participants were instructed to concentrate their attention on the musical excerpt as if they were trying to memorize it. While the musical excerpt was played for 60-seconds, a fixation cross was displayed at the screen center. Each of the musical excerpts was presented twice for a total of sixteen 60-second trials. The order of presentation avoided direct repetition of a given excerpt. The whole experiment was monitored using a custom version of OpenSesame developed by Oticon Medical (Omexp: Patou et al., 2019).

BEHAVIORAL INDICATORS

At the end of each trial, the participants were asked several questions (see table 1).

Questions	Possible answers	Behavioral indicator
Do you know this music?	Yes / No / Not sure	Familiarity
How much did your mind wander?	0-25% / 25-50% / 50-75% / 75-100% (of the time)	Mind Wandering

TABLE 1: QUESTIONS ASKED TO THE PARTICIPANTS AFTER EACH EXCERPT AND THE CORRESPONDING BEHAVIORAL INDICATOR.

For Familiarity we computed two categories across stimuli: Low and High familiarity. The Low familiarity category combines the answers No and Not sure while the High familiarity category is composed of the answer Yes. Importantly, because the musical excerpts were presented twice during the experiment, we only used the answer to the first presentation of each stimulus to avoid a repetition bias.

For Mind Wandering we computed two categories across participants: Low and High Mind Wandering, corresponding to participants for which the average mind wandering was lower or higher than the Grand mean, respectively.

Beside these behavioral indicators created from the questions, we also extracted two other indicators. The Presentation indicator, composed of two categories (i.e., First Presentation, Second Presentation and so on) and the Parts indicator also composed of two categories (i.e. First Part and Second Part) reflecting the first 30-second and the last 30-second of the musical excerpts. Importantly, these indicators allowed us to explore the effect of the acquired (short-term) knowledge as well as the impact of time on performances.

EEG DATA

EEG data were recorded using an ANT amplifier and a 21-electrode cap (10-20 system) with a sampling rate of 256 Hz. ICA was performed to remove line noise, eye-blinks and saccades artifacts and EEG data were digitally filtered between 1-9Hz and downsampled to 64Hz. Subsequently, they were cut into 30s epochs, resulting in 32 epochs.

AUDIO FEATURE EXTRACTION

Amplitude envelopes of the audio stimuli were obtained using the function `human_cochleagram` from the Python `Pycochleogram` package (<https://github.com/mcdermottLab/pycochleogram>). Each envelope was then digitally filtered between 1-9Hz, downsampled to 64Hz and cut into 30s long epochs in order to match EEG preprocessing.

STIMULUS RECONSTRUCTION

The stimulus reconstruction approach allows one to reconstruct an estimate of the envelope of the auditory stimulus using electrophysiological neural data via a linear reconstruction model (O'Sullivan et al., 2015). The reconstruction model is a temporal response function (see Lalor et al., 2009 for details) that maps neural data to the stimulus. A reconstruction score is computed, to estimate the reconstruction accuracy of the model, as the Pearson correlation between the reconstructed stimulus and the original stimulus. The higher the reconstruction score, the better the reconstruction accuracy (see Study One for more details on this statistical modeling).

STATISTICAL ANALYSIS

NORMALITY TESTING, OUTLIERS REJECTION AND EFFECT SIZES

When necessary, we assessed normality of the data distributions with a Shapiro-Wilk test. We tested for potential outliers using either a Grubbs test or a Tukey Fence for outliers. We assessed heteroscedasticity using either an F test or a Fligner test.

BEHAVIORAL DATA AND RECONSTRUCTION ACCURACY

In order to see whether the behavioral indicators were good predictors of the reconstruction accuracy we used linear mixed models (LMM) computed on data using R statistical software and the lmerTest package (Kuznetsova et al., 2014). To take into account the measures dependency, LMM included Subject and Stimulus as random effects. The normality of the distribution of model's residuals was checked using the qqPlot function of the EnvStats package (Millard, 2013).

RESULTS

Because the objective of this study was to investigate the relation between the behavioral indicators (familiarity, mind wandering) and the reconstruction accuracy of monodic trials, the reconstruction accuracy of monodic stimuli is briefly presented (see Study One for a full description of this result).

On average, the reconstruction performance of the monodic musical excerpts was 0.063 (± 0.08). and was significantly different from the null distribution ($t(40) = 14.32$, $p < 0.001$; Cohen's $d = 2.24$).

BEHAVIORAL INDICATORS AND PERFORMANCES

Importantly, participant responses were rather consistent with our initial stimulus categorizations of familiar versus unfamiliar music (Participants answered unfamiliar in 67% when the stimuli were unfamiliar and they answered familiar in 80% when the stimuli were familiar).

We wanted to explore the link between the behavioral indicators and the reconstruction accuracy of monodic stimuli. To do so, we first examined the main effects of the indicators. We found a significant positive main effect of Familiarity on the reconstruction accuracy (beta = 0.01, 95% CI [3.02e-03, 0.02], $t(1307) = 2.54$, $p = 0.011$; Std. beta = 0.16, 95% CI [0.04, 0.29]; figure 15 A) and the prediction model significantly predicted around 11% of the variance (conditional $R^2=0.11$). However, the part of explained variance related to the fixed effect alone was around 0.7% (marginal $R^2=6.52^{-3}$).

On the other hand, we found no main effect of the Mind wandering (p -value = 0.698) as well as of the Presentations (p -value = 0.872) reflecting the first and second presentation of the musical excerpts and Part indicators (p -value = 0.745) reflecting the first 30-second and the last 30-second of the musical excerpts.

Because we found a main effect of Familiarity, we then explored the interactions of Familiarity with the other indicators (figure 15 B).

We found a trend toward an interaction between the Familiarity and the Mind wandering indicators (beta = 0.02, 95% CI [0.04, 1.24e03], $t(1305) = 1.83$, $p = 0.067$; Std. beta = 0.22, 95% CI [0.45, 0.02]).

On the other hand, the Familiarity effect did not differ in the first and second part of the excerpts (Familiarity by Part interaction, p -value = 0.86) and was not affected by stimuli repetition (Familiarity by Presentation interaction, p -value = 0.55)

To further investigate this interaction, we computed the simple effects of the Familiarity on the reconstruction accuracy in the levels of Mind wandering (figure 15 B).

We found that the effect of Familiarity on the reconstruction accuracy behaved differently for the High Mind Wandering category (beta = 0.02, 95% CI [0.04, 8.64e-03], $t(1305) = 3.03$, $p = 0.003$; Std. beta = 0.30, 95% CI [0.50, 0.11]) and the Low Mind Wandering one (beta = -6.96e-03, 95% CI [-0.02, 5.19e-03], $t(1305) = -1.12$, $p = 0.262$; Std. beta = -0.09, 95% CI [-0.24, 0.06]).

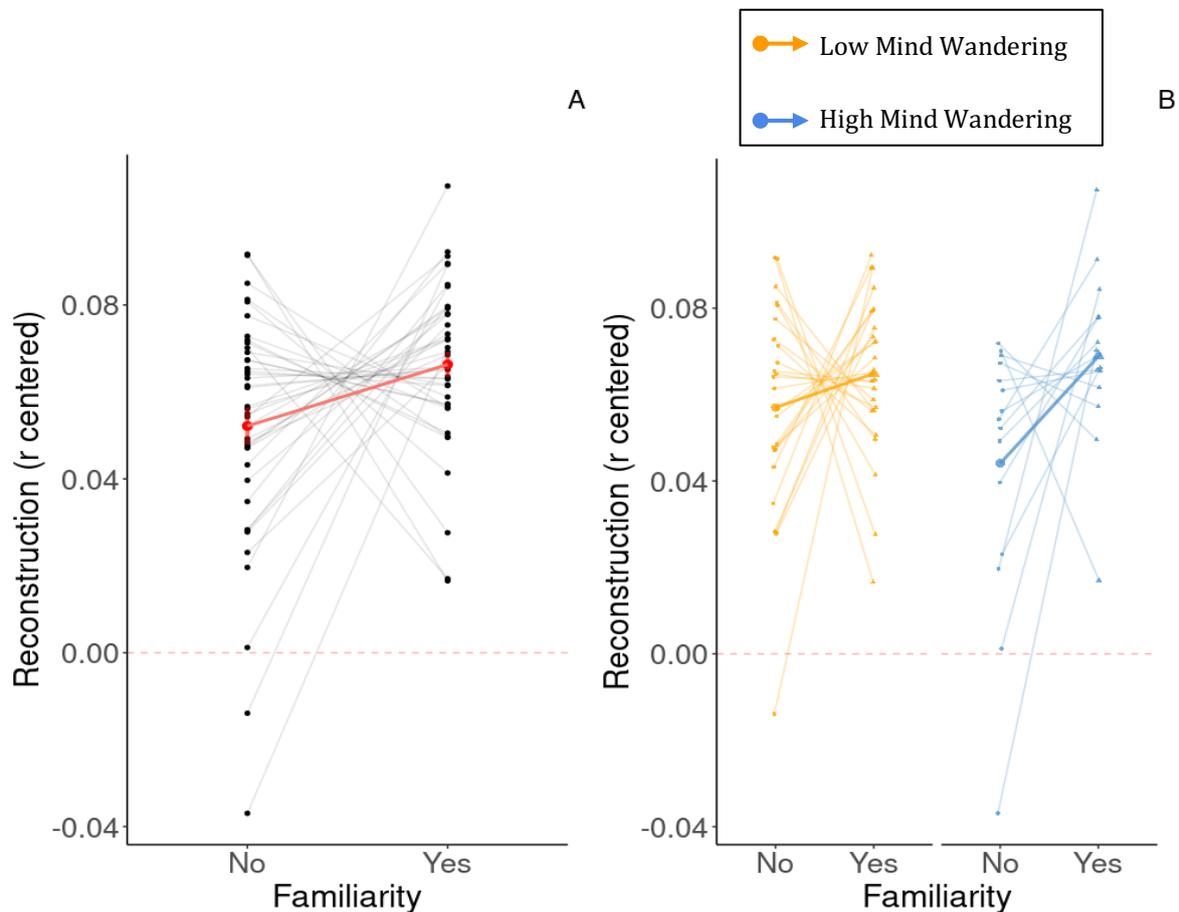


FIGURE 15: ILLUSTRATION OF THE MAIN EFFECT OF FAMILIARITY ON RECONSTRUCTION ACCURACY AND THE SIMPLE EFFECTS OF FAMILIARITY ON RECONSTRUCTION ACCURACY WITHIN THE MIND WANDERING INDICATORS' LEVELS. A) AVERAGE RECONSTRUCTION ACCURACY WHEN EXCERPTS WERE FAMILIAR AND UNFAMILIAR. AVERAGE RECONSTRUCTION WAS HIGHER FOR FAMILIAR THAN UNFAMILIAR EXCERPTS ($\beta = 0.01$, 95% CI [3.02E-03, 0.02], $P = 0.011$). B) AVERAGE RECONSTRUCTION WHEN EXCERPTS WERE FAMILIAR AND UNFAMILIAR AND WHEN MIND WANDERING WAS LOW OR HIGH. AVERAGE RECONSTRUCTION WAS HIGHER FOR FAMILIAR EXCERPTS WHEN MIND WANDERING WAS HIGH ($\beta = 0.02$, 95% CI [0.04, 8.64E-03], $P = 0.003$) BUT IT WAS SIMILAR FOR FAMILIAR AND UNFAMILIAR EXCERPTS WHEN MIND WANDERING WAS LOW ($\beta = -6.96E-03$, 95% CI [-0.02, 5.19E-03], $P = 0.262$). ERROR BARS INDICATE PLUS OR MINUS ONE STANDARD ERROR OF THE MEAN. LARGE POINTS INDICATE THE GROUP AVERAGES, SMALL POINTS INDICATE INDIVIDUAL DATA.

DISCUSSION

In this work, we hypothesized that reconstruction accuracy would be higher for familiar excerpts compared to less familiar ones. Our analysis confirmed this hypothesis and revealed a positive main effect of familiarity on reconstruction accuracy (Figure 1A). This result corroborates the findings of previous works showing an enhancement of cortical

tracking for familiar song utterances (Vanden Bosch der Nederlanden et al., 2022) and familiar natural music (Madsen et al., 2019; Weineck et al., 2022). The most probable reason why the cortical tracking of familiar stimuli is enhanced compared to unfamiliar ones is because they are more predictable as they elicited greater expectations (Almudena et al., 2020, Huron., 2006). Some recent results suggest that expectations may directly impact cortical tracking of music (Di Liberto et al., 2020a; Marion et al., 2021, Kern et al., 2022). And it has also been shown that stimulus predictability could have an influence on auditory scene analysis (e.g., Bendixen, 2014), although most of the studies investigating the influence of predictability on auditory scene analysis used paradigms in which concurrent sounds were present or where sound sequences contained deviant sounds.

Interestingly, one may argue that unfamiliar stimuli could have elicited greater neural responses, leading to better cortical tracking, because they are more prone to induced expectations violation than familiar ones, which could be in line with the results of Kumagai and colleagues (2017, 2018). However, extending the proposition of Weineck and colleagues (2022), it is likely that our unfamiliar stimuli did not elicit expectation violation because they were classical western music excerpts that were composed following the common rules of the genre. Thus, even if participants were not familiar with the particular excerpt, they were probably familiar with western music in some way. Consequently, it is probable that familiar music excerpts, similarly to more intelligible or familiar speech (Zoefel and VanRullen, 2015; Verschueren et al., 2020; Soni and Tata., 2021), induced better cortical tracking because expectations were higher compared to unfamiliar music and because unfamiliar music excerpts did not induce violation of expectations.

Besides an explanation in terms of better predictions for familiar stimuli, it is also possible that familiarity modulated cortical tracking because familiar music excerpts elicited a greater engagement/attention of the participants.

That being said, while the effect of familiarity was significant, its predictive power was rather weak (0.7% of the reconstruction accuracy) suggesting that, in our experiment, the familiarity with the musical excerpt probably does not massively impact its cortical

tracking. However, because most previous papers in the field do not report the predictive power, it remains difficult to assess the global impact of familiarity on cortical tracking.

In this work, we also assumed that the mind wandering quantity will negatively affect the reconstruction accuracy of monodic excerpts. Inconsistent with our hypothesis, we found that reconstruction accuracy was similar for all the participants, no matter if they experienced a low or a high quantity of mind wandering during the experiment. Although this result may sound surprising because of the detrimental effect of mind wandering on the task performance (Kam and Handy, 2013; Smallwood, 2013), we used a retrospective measurement of mind wandering requiring the participants to remember the number of times they catch their mind wandering during each trial which can suffer from forgetting and mental-aggregation biases (Kane et al., 2021).

But the most original result of this work is that the effect of familiarity was modulated by the global quantity of mind wandering (Figure 1B). Indeed, participants that experienced a high quantity of mind wandering during the experiment have higher reconstruction accuracy for familiar excerpts than for unfamiliar ones while participants that experienced a low quantity of mind wandering showed similar reconstruction accuracy for familiar and unfamiliar excerpts. This could indicate that when attention is mostly dedicated to the task at hand (i.e., listening to the musical excerpt) familiarity may not enhance the cortical tracking possibly because the total amount of attentional resources available can be devoted to actively listening to the excerpts, independently of whether it is familiar or unfamiliar. However, when attention is directed to unrelated thoughts rather than to the task at hand, leading the target musical excerpt to become “unattended”, the familiarity may compensate for this reduction of attentional resources devoted to the treatment of the musical excerpt. Then, familiar excerpts seem to be better represented at the cortical level when less attention is devoted to their processing. This idea could find support in some results suggesting that 1) although the underlying mechanism is still unclear, a cortical representation of the unattended sound can occur without direct attention (see for example O’Sullivan et al., 2015; Power et al., 2011; Hausfeld et al., 2018) and 2) familiar excerpts, because they are more predictable, are less computationally demanding (Southwell et al., 2017) which suggest that they can be processed with less attentional resources.

Interestingly, we found that the reconstruction accuracy was similar for the first and the second presentation of the musical excerpts which suggests no effect of short-term knowledge on cortical representation of monodic musical stimuli. This result is in line with the work of Di Liberto and colleagues (2020b) who showed no effect of repetition on cortical tracking of monodic synthetic music but not with the work of Madsen and collaborators (2019) who found a decrease in cortical tracking for familiar music across repetitions.

Overall, our results indicate that stimulus familiarity favors neural tracking. Importantly, this familiarity effect is present with long-term knowledge but not with short-term knowledge of the stimuli. Finally, this long-term familiarity effect on neural stimuli representations is modulated by attention and is maximal when participants are not well focused on the task at hand.

Further work is required to better understand how this effect of familiarity on cortical representation is driven by long-term knowledge about the musical excerpts.

GENERAL DISCUSSION

SUMMARY OF THE RESULTS

In this work we used a stimulus reconstruction approach to first explore whether it was possible to reconstruct and classify both monodic and dichotic natural polyphonic musical excerpts based on single-trial EEG recordings. We also investigated whether WM, sustained attention and attentional inhibition explain reconstruction and classification performance of the stimulus reconstruction-based AAD system. In addition, we explored whether musical expertise affects cognitive functions as well as neural tracking of musical stimuli by comparing non-musical and musical experts. Finally, we considered the influence of familiarity and mind wandering on the reconstruction performance of the stimulus reconstruction-based AAD system in the monodic condition only.

Our results demonstrated that it was possible to reconstruct and classify ecological polyphonic monodic and, although with lower accuracy, dichotic music stimuli based on single-trial EEG recordings.

We made the hypothesis that cognitive functions should influence the reconstruction and classification performance of the stimulus reconstruction-based AAD system in both conditions. For monodic stimuli we found no effect of the selected cognitive functions over reconstruction and classification performance which invalidated our hypothesis. For dichotic stimuli however, attentional inhibition did explain around 10% of the reconstruction performance and around 8% of the classification performance which partially confirmed our hypothesis because no other cognitive function was a significant predictor.

We also hypothesized that musical expertise would influence both the stimulus reconstruction-based AAD system's performance in monodic and dichotic conditions and the executive functions. Unlike what we expected, we found no effect of musical expertise on reconstruction and classification performances for monodic nor dichotic stimuli. In addition, we found no difference in terms of executive functions between participants with high and low musical expertise.

Finally, we expected that, in the monodic condition, the familiarity with the musical excerpt would positively impact its reconstruction while the mind wandering propensity

would have a negative effect. Our analysis partially confirmed our hypothesis: the effect of familiarity was significant, although its predictive power was weak (0.8% of the reconstruction accuracy). This effect was modulated by the level of mind wandering of the participants.

The general discussion is structured in two parts.

In the first part, we discuss the results in a more general perspective. First, we explain the rationale behind the use of ecological music stimuli. Then, we provide elements about the potential issues of our protocol. Finally, we propose some reflections around the potential nature and timing of the attentional inhibition mechanism as well as on the reasons why the role of working memory, sustained attention and musical expertise remains unclear in challenging auditory scenes.

In the second part, we discuss several perspectives for this work starting with the relevance, as well as the current limitations, of the EEG-AAD framework for the next generation of cochlear implants. Then, we discuss the role that our results may play in contributing to the challenge of performance variability in AAD systems. We also propose some evidence for the benefit of specific cognitive training to improve AAD systems. Then, we discuss the relevance of using music stimuli for such a training. Finally, we consider some plausible applications of AAD systems in other domains.

ON THE USE OF MUSIC STIMULI

In this work, we decided to use real music for two main reasons. The first reason is linked to the importance of generalizing scientific results. Indeed, the vast majority of the studies having used a stimulus reconstruction-based AAD system generally used naturalistic speech stimuli, such as stories (e.g., O'Sullivan et al., 2015; Mirkovic et al., 2015; Schäfer et al., 2018; Fiedler et al., 2017). This makes sense because the stimulus reconstruction-based AAD allows one to investigate the neural processing timeline of speech (Crosse et al., 2021) in monodic conditions as well as the impact of attention in natural cocktail party scenarios in dichotic conditions. Comparing speech and music allows a better understanding of the neural basis underlying a given process as well as a more precise definition of the computation itself (Besson and Schön, 2001). Then, if one wants to explore the processing timeline of natural music as well as the impact of attentional processes in more natural settings, the first and most fundamental step is to be sure that

the stimulus reconstruction-based AAD can be used with natural music. Our results confirm the first attempts that have been done in this direction (e.g., Cantisani et al., 2019; Di Liberto et al., 2020b; Hausfeld et al., 2021) by showing that the stimulus reconstruction-based AAD can be used with polyphonic natural music stimuli in both monodic and dichotic conditions. This is important because it suggests that this approach is appropriate to explore the processing of natural polyphonic music.

The second reason is because music can be polyphonic at multiple levels, which allows one to create more or less complex auditory scenes. Here, we decided to use piano excerpts because we supposed that using pairs of polyphonic instruments in the dichotic condition would allow us to create more neutral auditory scenes in terms of attentional capture. If we had used pairs of two distinct monophonic instruments, participants could have been attracted more by one instrument or the other and if we had used two distinct polyphonic excerpts (e.g., symphonies) it is possible that participants could have failed because of auditory scene complexity. Overall, the choice of using natural piano excerpts, emerges from a trade-off between keeping polyphony, reducing attentional capture and making a task that was difficult enough but not impossible. In addition, we chose the piano because this instrument should elicit a rather good cortical representation of envelope fluctuations due to its faster attack compared to other traditional instruments such as violin or cello. Although some studies found good reconstruction accuracy ($r = \sim .15$) using bassoon and cello for instance (Hausfeld et al., 2021), they used synthetic music pieces. However, our choice of using piano may be challenged by the work of Cantisani and colleagues (2019) which indicates that reconstruction accuracy of other instruments (e.g., cello, bass, guitar, horn) may be equivalent using a similar procedure and real music recordings.

ON THE IMPACT OF MUSIC STIMULI ON ATTENTIONAL PROCESSES

IMPORTANT EXCERPTS DYNAMICS MAY HAVE PROMOTED ATTENTIONAL SWITCHING

One can argue that using ecological stimuli, which involve highly variable dynamics (as an expressive element), may have favored participants to switch attention between the two excerpts. This could partly explain the rather low reconstruction accuracy of attended musical excerpts ($r = \sim .03$). One could wonder whether this implies that the attentional inhibition mechanism was not sufficiently performant to prevent the salient

event from capturing attention. It is not easy to provide a clear answer to this question based on our data. However, one may speculate and propose that it is maybe impossible to completely suppress the irrelevant stimuli (the concurrent stream) and therefore, events that exceed a certain threshold will automatically capture attention in a bottom-up fashion. The fact that the irrelevant stimuli cannot be completely inhibited could have been inherited from an archaic survival mechanism. Indeed, such a mechanism truly has an interest for evolutionary purposes because it could act as an alerting system that automatically orientates attention to an irrelevant salient event, even when attention is fully devoted to a relevant stimulus. This explanation could be in line with the fact that 1) concurrent streams are generally represented at a cortical level even without direct attention (Hausfeld et al., 2018) and 2) the evidence suggesting that unpredictable irrelevant sounds may be more difficult to inhibit (Devergie et al., 2010; Rimmele et al., 2012).

ABSENCE OF BOTTOM-UP CUES MAY HAVE FAVORED ATTENTIONAL INHIBITION

In the dichotic condition, we created a complex auditory scene in which it was rather difficult to segregate the two music excerpts solely based on the bottom-up cues because the stimuli composing the pairs were matched in volume, rhythm and timbre. Nevertheless, we decided to separate the streams in the spatial dimension because otherwise it would have been too difficult to both segregate and follow them through time. It results that the task was difficult but not impossible, as orally reported by the participants. That being said, this peculiar design has probably led to a situation in which attentional inhibition was “forced” to come into play. Indeed, while the segregation of the two streams could have been achieved solely using spatial bottom cues (i.e., the location) and therefore possibly without inhibition, it is likely that the effective streaming of the target music excerpt cannot have been achieved without inhibiting the concurrent one. Indeed, the energetic masking created by the spectro-temporal overlapping between the target and the distractor excerpts leads to a small perceptual distance and similar temporal coherences between the two streams. Therefore, participants were required to inhibit the concurrent stream because bottom-up cues cannot help them to follow the target stream through time. This is interesting because on the one hand, this design allows us to see that attentional inhibition seems to be important for an efficient cortical tracking of music in complex auditory situations and that it predicts an important part of

the variance of the performance of a stimulus reconstruction-based AAD system. But on the other hand, the generalization of our results remains difficult because it is hard to say that attentional inhibition will be as important in all kinds of auditory scenes. Especially in auditory scenes where bottom-up cues are available and may help the segregation as well as the following of streams in time such as situations where energetic masking between the target and concurrent streams is low (e.g., polyphonic music with two different instruments, target speech with concurrent music background). Moreover, it is unclear whether the impact of attentional inhibition will be the same with different stimuli types such as speech. Further works should use the same approach to see to what extent attentional inhibition is involved in auditory streaming of speech but also if other cognitive functions are good predictors of the streaming ability in speech cocktail parties.

ON THE POSSIBLE NATURE AND TIME COURSE OF THE ATTENTIONAL INHIBITION MECHANISM

As previously mentioned, we found that the ability to resolve stimulus-based conflict in an auditory Stroop task predicted a substantial amount of the reconstruction performance of a stimulus reconstruction-based AAD system. Thus, it suggests that the ability to resolve stimulus-based conflict may be important to be able to concentrate on a specific target stream and follow it through time, in a challenging auditory environment. In our specific case, one possibility is that this ability to resolve stimulus-based conflict comes into play as an attentional facilitation mechanism that enhances the features representations of the target stimulus. This claim would be in line with previous propositions of an enhancement of the attended stream (e.g., Mesgarani and Chang, 2012; O'Sullivan et al., 2015; Zion Golumbic et al., 2013; Hausfeld et al., 2018) and it could also fit with hypotheses about neural strategies for conflict resolution (Desimone and Duncan, 1995; Kastner and Ungerleider, 2000; Egner and Hirsch, 2005; Egner et al., 2007). However, another, although nonexclusive, explanation could be that this ability to resolve stimulus-based conflict come into play as an inhibitory mechanism, reducing the concurrent (goal-irrelevant) stimulus features representations at the cortical level (or even earlier), which is consistent with the results of studies showing both facilitatory and inhibitory mechanism of competing sounds (e.g., Bidet-Caulet et al., 2010; Chait et al., 2010). Therefore, it is possible that two mechanisms, a facilitatory one and an inhibitory one, cooperate to facilitate the processing and the streaming of the to-be-attended stream

in challenging auditory situations. Although it has been shown that these two mechanisms have different timing in simple auditory scenes (e.g., Bidet-Caulet et al., 2010; Degerman et al., 2008; Chait et al., 2010), the facilitatory one starting earlier (around 100 - 150ms) than the inhibitory one (around 200 - 250ms), it is still unclear whether this dissociation remains in complex auditory scenes and whether the timings are the same. Previous studies using stimulus reconstruction have computed single lags models that represent the reconstruction accuracy of the attended and/or unattended streams at different time-lags (moments) instead of multi-lag models (i.e., the procedure we used in this work) that provide a single value indicating overall tracking performance. This procedure is useful to explore at which stages of the auditory processing attention modulation of cortical tracking occurs and therefore to obtain a finer view of the time course of the attentional mechanisms at work during complex listening scenarios. Using this procedure, several studies found that, for speech, higher reconstruction accuracy appears in the interval of 130-250ms post-stimulus (e.g., O'Sullivan et al., 2015; Mirkovic et al., 2015, Simon et al., 2022). While this time window appears rather similar to the one obtained in previous studies using ERP (e.g., Bidet-Caulet et al., 2010), it remains hard to say that inside this window a facilitatory and an inhibitory mechanism are at work. In addition, very recent studies found that this time window of optimal cortical representation was later for music stimuli (Hausfeld et al., 2018, 2021; Simon et al., 2022) which indicates that the attention dynamic could be different for music and speech.

That being said, if we suppose that the inhibitory mechanism will reduce the concurrent stream's cortical representations, we can suppose that its reconstruction accuracy will be low. Therefore, a possible way to obtain information about the timing of the inhibitory mechanism could be to look at the time-lags (i.e., the post-stimulus moments) where the reconstruction accuracy of the unattended stream is the lowest (see figure 16 for an illustration). Another way could be to look at the strength of the relation between attentional inhibition (as well as other cognitive functions) and the reconstruction accuracy at each time lag. By doing so, it would be possible to see at which moment of the stimulus processing the correlation between attentional inhibition and the reconstruction accuracy is maximal.

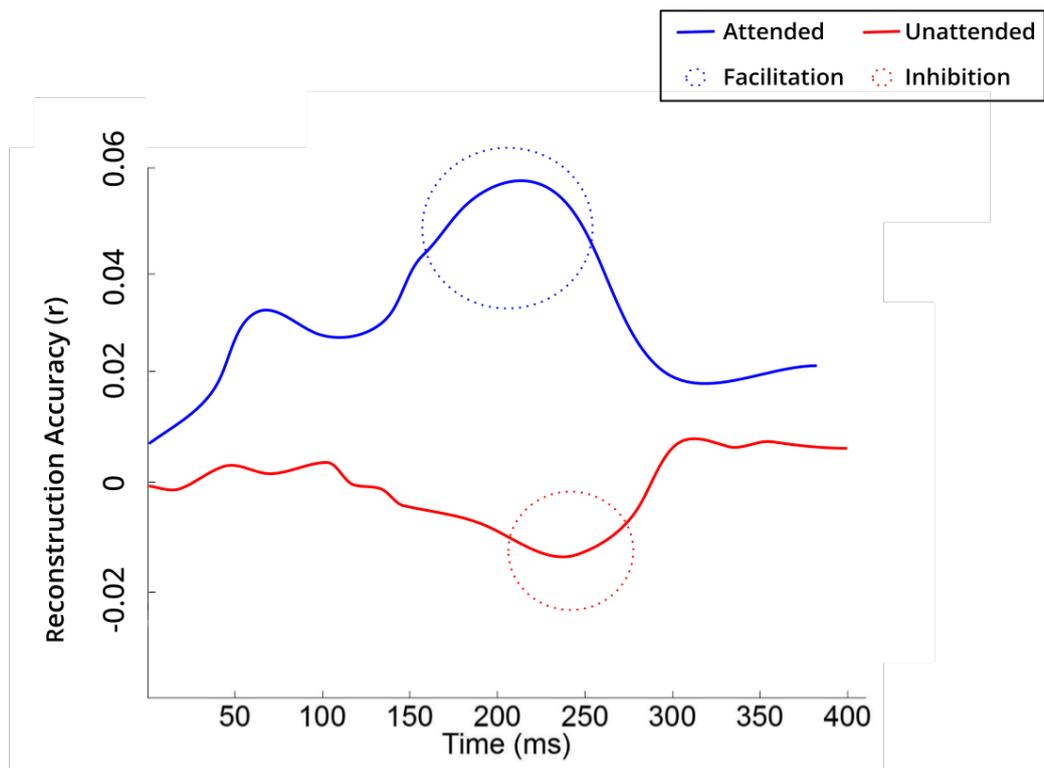


FIGURE 16: THIS FIGURE ILLUSTRATES THE POTENTIAL TIMING OF THE FACILITATION AND INHIBITION MECHANISMS. IN THIS ILLUSTRATION, THE ATTENTIONAL FACILITATION PEAKS AROUND 220 MS POST STIMULUS AS INDICATED BY THE HIGHEST RECONSTRUCTION ACCURACY OF THE ATTENDED STREAM, IN THE BLUE DASHED CIRCLE. THE ATTENTIONAL INHIBITION PEAKS AROUND 250 MS POST STIMULUS AS INDICATED BY THE LOWEST RECONSTRUCTION ACCURACY OF THE UNATTENDED STREAM, IN THE RED DASHED CIRCLE.

NOTE THAT THE FIGURE DOES NOT REPRESENT REAL DATA BUT SYNTHETIC DATA CREATED FOR ILLUSTRATIVE PURPOSES ONLY.

ON THE IMPORTANCE OF WORKING MEMORY AND SUSTAINED ATTENTION FOR AUDITORY STREAMING IN CHALLENGING ENVIRONMENTS

Contrary to attentional inhibition, we found that working memory capacity and sustained attention do not predict the performance of the stimulus reconstruction-based AAD system. The first interpretation suggests that working memory and sustained attention capacity are not important for auditory streaming in challenging situations while attentional inhibition is. However, the interpretation of this result is not so straightforward because 1) working memory and sustained attention have been linked to performance in tasks requiring auditory scene analysis (e.g., Akeroyd, 2008; Sorqvist and Ronnberg, 2012; Escobar et al., 2020) as well as in dichotic listening tasks (Colflesh and Conway, 2007; James et al., 2014) and 2) it is difficult to disentangle working memory, sustained attention and attentional inhibition. Indeed, in many models of

working memory, the ability to inhibit distractors is an important mechanism allowing an efficient maintenance of information (Baddeley, 1996; Cowan, 2001; Engle and Kane, 2004; Barrouillet et al., 2004). And inhibition of distractors is also crucial to be able to sustain attention on a specific task. Thus, it is possible that the cognitive tests we used were not appropriate because the manifestation of working memory and sustained attention in the context of auditory streaming in a challenging auditory scene is too different from what we were measuring with the selected cognitive tests. In addition, this result is also consistent with the idea that attentional inhibition could be the mechanisms whereby working memory and sustained attention are involved in complex auditory scene analysis.

ON THE ROLE OF MUSICAL EXPERTISE IN AUDITORY STREAMING IN CHALLENGING AUDITORY SITUATIONS

Although one can find inconsistent results in the literature about the impact of musical expertise on executive functions, attention and auditory scene analysis, there are several cues suggesting that musicians could be better than non-musicians. Indeed, some works have shown better performance of musicians in cognitive functions involved in auditory scene analysis tasks, including attentional inhibition (e.g., Bialystok et al., 2009; Kaganovich et al., 2013; Sharma et al., 2019), working memory (e.g., Clayton et al., 2016; Suarez et al. 2016; Zuk et al., 2014, see Talami et al., 2017 for a review) and attention (Roman-Caballero et al., 2021; Strait and Kraus, 2011; Pallesen et al., 2010). Other works have found evidence about a musician's advantage in tasks requiring auditory scene analysis such as stream segregation (Zendel and Alain, 2009; François et al. 2014), speech in noise listening (Parbery-Clark et al., 2009a; Coffey et al., 2017) or dichotic listening (Barbosa-Luiz et al., 2021). Finally, several studies demonstrated a better cortical tracking of speech and music stimuli for musicians compared to non-musicians (Di Liberto et al., 2020b; Doelling and Poeppel, 2015; Puschmann et al., 2019).

However, while there is a trend toward a better reconstruction performance of musicians compared to non-musicians in the most challenging condition (i.e., the dichotic listening), it is interesting to note that we do not observe a salient effect of musical expertise in a task in which one can reasonably think that musicians would outperform non musicians, taking into account the previous findings. Several non-exclusive explanations may be discussed here.

First, the most plausible reason may rely on the fact that the two music excerpts were piano pieces that were matched in rhythm and timbre. This implies very small differences in the spectral and temporal features of the two excerpts. The point is that these bottom-up cues are thought to be important to segregate streams. And in addition, a musician's advantage has mainly been reported for pitch discrimination (e.g., Micheyl et al., 2006; Magne et al., 2006; Schön et al., 2004) and timing perception (e.g., Rammsayer and Altenmüller, 2006; Danielsen et al., 2022) and the "superiority" of musicians in speech in noise listening may be due, at least in part, to a finer frequency discrimination ability (e.g., Parbery-Clark et al., 2009b; Fuller et al., 2014). All these elements suggest that, in conditions where musicians are less able to use spectral and temporal cues in stream segregation, their advantage may have been reduced.

The second explanation is related to the importance of inhibiting irrelevant streams in our task. Indeed, as mentioned earlier, attentional inhibition is key in our task because the unique stimulus feature that can help them segregate the two streams is their spatial location. However, while studies exploring the difference between musicians and non-musicians in inhibitory control have, rather consistently, reported a better response inhibition in musicians (e.g., Moreno et al., 2011, 2014; Moussard et al., 2016), it is way less clear that attentional inhibition is enhanced in musicians compared to non-musicians (e.g., Bialystok and DePape, 2009; Sharma et al., 2019 for congruent results and Schellenberg, 2011; Vanden Bosch der Nederlanden et al., 2020; Clayton et al., 2016 for incongruent results). Therefore, it is possible that we did not find any musician's advantage in our task because 1) response inhibition was not engaged and 2) musicians do not have better attentional inhibition than non-musicians.

A third reason, although more speculative, may be that active listening to polyphonic music could enhance attentional inhibition in challenging situations as much as musical training. Indeed, one may reasonably think that people who are actively and frequently listening to polyphonic music are in fact training their ability to segregate streams. In polyphonic music, one can choose to integrate the different melodic lines to listen to a unique percept (i.e., "the whole thing"). But a second option is to selectively listen to a specific melodic line (i.e., the violin part in a symphony) which requires attentional inhibition of the other melodic lines.

Therefore, it could be that simply listening to polyphonic music is challenging enough to make the auditory system an expert, thus somewhat minimizing the further impact of active music making.

PERSPECTIVES

This first evidence that attentional inhibition, as assessed by a behavioral cognitive task, may have a predictive power of the performance of a stimulus reconstruction-based AAD system raises the question of the possibility to predict the performance of such a system based on cognitive measurement. It is particularly interesting from the perspective of Neuro-steered hearing aids as well as from a cognitive training standpoint. Consequently, we will further discuss these points in the following sections.

TOWARDS A NEW GENERATION OF COCHLEAR IMPLANTS?

For decades now, researchers have tried to improve cochlear implants by changing the number of electrodes, their pulse strategies or by creating new background separation algorithms (see for example Goehring et al., 2019). Generally, these techniques have improved the performance of the cochlear implants in moderately simple environments, meaning, when there is one source of interest and relatively stationary background noise. However, it does not solve the issue when the acoustic scene becomes more complex, especially when several stimuli in the auditory scene are of the same type (e.g., speech, music). However, since the recent development of the Auditory Attention Detection framework, it is plausible to imagine a neuro-steered cochlear implant that could overcome these limitations.

An easy way to consider a neuro-steered cochlear implant is to think about a system composed of two modules, one responsible for auditory attention detection (AAD module) and another one which is the cochlear implant *per se*. The AAD module would be in charge of providing information about the source of interest estimated via real-time neural activity measurement to the cochlear implant which would selectively amplify the user's source of interest, in a complex acoustic environment. However, as mentioned by Geirnaert and colleagues (2021), developing such a system is not straightforward because it will require a lot of "technical" progress. For instance, in day-to-day settings, the AAD algorithm should detect the source of interest in a real-time or near real-time fashion which is still not the case, even for the best algorithms, especially in contexts

where the user is free to switch attention between several sources, where reverberation is present and the auditory environment is constantly evolving. In addition, in a real-life environment, the EEG system cannot be an expensive, heavy system composed of a cap with a lot of electrodes that necessitate conductance gel and electric interference-free environment to properly work. Instead, it has to be easy to use, light, small, dry and resistant to electric interference. And there are many other problems that need to be solved before such a system will be available for a large audience (Geirnaert et al., 2021). Hopefully, quite a lot of efforts and resources have been dedicated to solving these issues since the last five years. Nevertheless, the important issue of the variability of the AAD systems' performance that may tarnish the success of the neuro-steered cochlear implants remains largely ignored.

SOLVING A PART OF THE VARIABILITY PROBLEM OF AAD PERFORMANCE WITH COGNITIVE FUNCTION BASED PREDICTIONS

The problem of the variability in the performance of the AAD systems is a big challenge for the development of neuro-steered hearing aids. Actually, it is still unclear why AAD systems are working very well for some individuals but much less for others. One source of variability is probably related to physiological factors such as the cranial thickness, the cortical orientation with respect to the skull or the signal to noise ratio of neural data that could be different for each individual. However, another source of variability may be related to higher order factors such as the effective ability to perform the task, meaning being able to efficiently concentrate attention on the source of interest. In this work we supposed, based on the literature, that the ability to concentrate attention on the source of interest depends on several cognitive functions (working memory, attentional inhibition and sustained attention). We went further, by making the hypothesis that it could be possible that poor cognitive functions may impact the performance of an AAD system because these cognitive functions influence how an individual will be able to concentrate attention to a target auditory stream. Our results demonstrated, for the first time, that attentional inhibition predicts a substantial amount of the performance of an AAD system (around 10%). This means that attentional inhibition is probably important in challenging auditory scenes. But more interestingly, it also suggests that by measuring attentional inhibition with a rather short, computerized and accessible cognitive test (i.e., online Stroop test) it could be possible to predict the performance of the AAD system. For

adults and elderly, in the context of a neuro-steered hearing aid, one can think about a pre-implantation monitoring of the performance based on cognitive functions testing. This pre-implantation performance monitoring could take the form of a battery of cognitive tests including attentional inhibition and potentially other cognitive functions (further works are needed to identify the most relevant cognitive predictors). For children however, a pre-implantation monitoring will be challenging before surgery as it occurs around 18 months. However, it is still possible to assess the performance of the neuro-steered hearing aid later in order to orient toward cognitive interventions or speech therapy if needed.

That being said, in this work we used an auditory version of the Stroop test which obviously means that individuals suffering from hearing loss will not be able to correctly perform the test. This is an important point considering that it is still unclear whether attentional inhibition is amodal or not. Therefore, it is unknown whether the predictive power of attentional inhibition would have been the same if we had used a visual Stroop test. Further work is thus needed to investigate this specific point.

In addition, the degrees of hearing loss as well as the moment of their emergence are manifold and may differently impact cognitive functioning. For instance, it has been hypothesized that early auditory deprivation due to hearing loss may lead to cognitive dysfunctions (Lieu et al., 2020), which could contribute to poor auditory scene analysis abilities. Consequently, it remains unclear to what extent our results, obtained with normal hearing individuals, can be generalized to individuals suffering from hearing loss.

Overall, these results offer promising evidence about the possibility to predict the performance of a neuro-steered hearing aid solely based on common cognitive tests. Nevertheless, research remains necessary to be certain that this methodology could be used with persons suffering from hearing loss.

IMPROVING AAD SYSTEM'S PERFORMANCE WITH SPECIFIC COGNITIVE FUNCTIONS-BASED TRAINING

Beside the predictive power of specific cognitive functions on AAD system's performance *per se*, our results also suggest that specific cognitive functions' training could enhance the performance of an AAD system. Indeed, our results revealed that attentional inhibition seems to be important for auditory streaming in challenging auditory

situations. Therefore, training attentional inhibition could improve the ability to inhibit concurrent auditory sources and subsequently improve the cortical tracking of the target one. For patients suffering from hearing loss, and specifically cochlear implant users, it is interesting in the context of neuro-steered hearing aids. Indeed, cochlear implant users are less able to rely on bottom-up cues to segregate an auditory scene, due to the poor spectral resolution of the cochlear implant. Therefore, enhancing their ability to inhibit irrelevant sounds may be highly relevant to help them perform better auditory scene analysis. In addition, supposing that improving the ability to inhibit irrelevant sounds could be a lever to improve the detection of the target auditory source and subsequently the performance of the neuro-steered hearing aid, cognitive training may be highly relevant for people suffering from hearing loss. Such a training could be even more relevant because the underlying functions of auditory scene analysis (e.g., attentional inhibition) are generally not trained in traditional therapeutic interventions (Roman et al., 2016).

That being said, a training of attentional inhibition could also be beneficial for normal hearing individuals that experienced difficulties in complex auditory situations as well as for children, to enhance the development of this cognitive function in capitalizing on the critical window and their higher brain plasticity. In addition, it is also possible that attentional inhibition training could benefit individuals suffering from attention-deficit/hyperactivity disorder (ADHD) because executive function deficits are thought to be a central feature in this pathology (see Lambez et al., 2020 for an interesting discussion about nonpharmacologic therapies for ADHD). In fact, such an attempt was made by Cho and colleagues (2002) who developed an attention enhancement system for ADHD children using EEG biofeedback and a virtual classroom environment. They showed that it was possible to use a brain computer interface (BCI) to enhance attention in children with ADHD in a school context.

Overall, it seems that attentional inhibition training may help a variety of individuals including people suffering from severe hearing loss, patients with attentional disorders but also healthy children and adults.

IS MUSIC-BASED TRAINING THE BEST OPTION?

In our work, we found no effect of musical expertise on the performance of our AAD system which suggests that musicians were not better than non-musicians to follow a target stream in our challenging auditory environment. In addition, we found no effect of musical expertise on any of the cognitive functions we were interested in. As we proposed in the previous section, it is possible that we did not find any effect of musical expertise here because our participants were “music lovers” that trained their cognitive functions and more particularly attentional inhibition without even knowing it by listening to polyphonic music. Anyhow, it is still possible that musical expertise does not enhance auditory scene analysis ability in very challenging environments because these situations are too specific and too different from what a musician is used to. Moreover, the majority of musician's training rests on practicing their own instrument, often alone which probably does not extensively engage attentional inhibition. Overall, it means that the door remains open to cognitive training that uses music material, which could have several advantages (Torppa and Huotilainen, 2019; Pesnot Lerousseau et al., 2020). The first advantage is the rather recreational aspect of music compared to speech, particularly for children but also for adults. The second advantage is the flexible structure of music that can vary at different scales. For instance, one may change the rhythmic structure (note rate), the phonic structure (monophonic, polyphonic), the spectral structure (timbre, pitch) of a music piece or even the location of the instruments. The interesting point is that it is possible to mix these different structure variations to create more or less complex auditory scenes. In that sense, it is possible to create training that is very well adapted to the user, with progressive difficulty levels and tackling specific processes. In addition, polyphonic music allows for both segregation and integration of the different melodic lines which could be an interesting lever to modulate the complexity of an auditory scene.

On the other hand, the counterpart of using music material is that we do not know for sure whether the benefit of the training will extend to other situations with other types of material (a cocktail party with target and background speech for instance). In addition, for cochlear implant users, using music stimuli may be difficult because of the poor spectral resolution of the device. That being said, one can think about other training using

a mixture of music and speech as well speech in speech stimuli. The good question to ask remaining: *what is the most appropriate training for each individual?*

In the end, an AAD system could be a useful tool to estimate the impact of a specific cognitive training on the auditory scene analysis abilities as well on the processing of natural sounds such as speech (Gillis et al., 2022) and, of course, music.

OTHER APPLICATIONS OF AUDITORY ATTENTION DETECTION

Aside from the application of AAD in the development of a new generation of cochlear implants and cognitive training, one can also imagine other applications of this technique in the world of Brain-Computer Interface (BCI).

In the field of art for instance, several attempts have been made to bridge EEG and BCI since the 1970s (Vidal, 1973; Rosenboom, 1977; Williams and Miranda, 2018). More recently, works have been done to develop systems to control an instrument (Arslan et al., 2005) or to generate melodies with brain signals (wu et al., 2010; Miranda et al., 2011) to name a few. In this sense, there is a place for AAD-BCI systems to create new kinds of art performances in which brain activity induced by auditory attention could be used to modulate different sound sources. This could be of particular interest in an immersive listening structure composed of multiple loudspeakers (Pascal, 2020). Such a device would allow the user to select a specific sound source and modify its loudness, spatial location or motion. In such a setup, the AAD module monitors in real-time the attended source and provides information about the source of interest to the BCI module. This second module is responsible for analyzing the intentions of the user, translating them into command and controlling an external device. To do so, the BCI module classifies among several classes of neural activity induced by different cognitive processes (e.g., imaging a movement of the attended source). Once the user's intention has been detected, the pBCI module translates it into commands that correspond to a particular parameter's modification (e.g., moving the attended source from the upper central loudspeaker to the bottom central one) and sends them to an external device.

Beside arts, one can think about other futuristic applications for AAD in several distinct domains. For instance, in the entertainment field where it could be possible to develop "auditory games" in which players, equipped with light AAD-pBCI systems, confront each other in musical battles using their auditory attention. In addition to being fun, this kind

of game could be interesting to develop cognitive abilities that underlie auditory scene analysis such as WM, executive control or selective attention, even if it is not its main purpose. Furthermore, such a game could be adapted to a solo or a multiplayer environment. AAD-BCI systems could also find applications in the field of domotics. Indeed, a wearable AAD-BCI system could be useful, in situations where ambient noise is varying constantly (e.g., in a living room), to monitor and adapt in real-time the loudness of the attended sound source (TV, hifi system, home phone, etc.).

CONCLUSION

In this work, we were interested in better understanding to what extent cognitive factors could impact the performance of a reconstruction stimulus-based AAD system, in a music listening context. Our results show the feasibility of using the reconstruction stimulus-based AAD approach with ecological polyphonic musical stimuli and point to specific cognitive mechanisms being at work during dichotic musical listening in normal hearing listeners. In addition, they suggest that training specific cognitive functions could be useful to improve AAD systems and by extension, a next-generation of hearing aids.

Although we provide some hints about the potential reasons why the performance of AAD systems is so variable, further effort is necessary to fully appreciate the importance of cognition on AAD systems. One point seems highly relevant to me: in the AAD domain, one should not overlook cognition, because even the best AAD algorithm will be strongly affected by the cognitive state and skills of each individual.

BEHIND THE SCENES...



Over the course of this 3 years adventure, we faced many troubles that had more or less incidence on its smooth progress. Rather than providing an exhaustive list of these problems, which, in my opinion, would have been extremely boring, I decided to present a chronological overview of the whole adventure, where the important (bad or good) events are summarized, year by year.

Of course, you are free to skip this part but be aware that what comes next is the reference part. And this is definitely more boring!

September 2019 - September 2020

Year 1: The great beginning

A lot of events took place during the first year of this adventure. Of course, I could have started by saying that, I started my PhD!! But I suspect that you already know this information... Anyways, I think that one of the most important events that occurred during this first year was probably the beginning of the codirection, in November, because it has reshaped everything. From this moment, we had to change the thesis subject, moving from a pure algorithmic/brain-computer interface topic (i.e., real-time detection of auditory attention during music listening) to a more cognitive-science one (i.e. the influence of cognitive functions and musical expertise on a music-based AAD's system). Of course, changing the subject after the beginning of the thesis was not an easy task, especially because we had several constraints imposed by the different actors of this project, and it took us a lot of time (by a lot of time I really mean a lot of time), a loot of brainstorming and a looot of reading to finally find elements that were interesting and sufficiently innovative. The good aspect being that we were able to accentuate the cognitive side of the thesis.

In the middle of March, the second most important event of this first year happened. Covid-19 comes out, forcing us to stay home and slowing down everything. To make a long story short, this was a rather dark period for (my) science. We lost a considerable amount of time, especially because this small guy destroyed all our chances to test our experimental set-up in real conditions for months! To be honest, it was somewhat hard for me to work efficiently during the lockdown, probably because I was not mentally and logistically prepared to handle such a situation. However, on the “positive” side of the thing, it has opened our eyes about the possibilities for us to use online experiments to

both launch pilot studies to test our ideas and to measure cognitive functions to avoid fatigue for our participants. Until the month of September 2020, we worked on the experimental set-up, essentially on how to synchronize EEG and pupillometry (*oh yes, I did not mention that we chose to add pupillometry in our experimental protocol. Was it a good idea? Well, I think I am still not 100% sure about that*) and refine the online cognitive experiments. Finally, this first year ends up with no major breakthrough but a lot of interesting ideas for the rest of the adventure.

I certainly have to mention that, at the very end of August, I attempted the ILCB summer school, my first real scientific event as a PhD student. For sure it was an amazing experience!

September 2020 - September 2021

Year 2: Entering Macgyver mode

Frankly, it would be a lie to say that this second year was a long calm river. On the scientific part, I think it was a rather good year. Of course, we have had failures. Between September and December 2020, we spent time designing an experiment to measure what we named the “attentional window”, in which participants were instructed to concentrate their attention on the left or the right stimulus and report each time they switched their attention. The idea was promising but it failed, for many reasons that I will not describe here. Nevertheless, this failure helps us to become aware of several parameters that we will have to take into account to design a new experiment. During the first month of 2021 we designed a new experimental protocol based on the ashes of the previous one (which will turn out to be the protocol described in the thesis) which would allow us to investigate the influence of cognitive function on the performance of an AAD system. Importantly, during this period, we also wrote a mini-review article about AAD algorithms and their possible applications in the world of non-clinical BCIs, which was published in *Frontiers in Computer Science* in April 2021! What an achievement!

But the true first major event of this year happened at the end of April 2021 with the beginning of the experiment phase! It was really an exciting period for me but in reality, nothing was ready. We discovered issues with participant answers recording in the online experiments and we realized that the EEG and the pupillometry recording were not synchronized at all because of the software tagging. Solving the first problem was an

easy task but after we strived for several weeks to find solutions to keep software tagging, it was decided to use hardware tagging to synchronize the two devices. And the miracle of the Holy StimTrack happened! The experimental set-up was ready to get back in business and at the beginning of the summer the experiments started again.

However, at the very end of the summer, after 5 participants, something terrible happened... During an innocent postal exchange, we lost the StimTrack! Consequently, this second year ends up with a mixed feeling. On the one hand I was delighted that our protocol was functional and ready to use. On the other hand, I was frustrated that I cannot effectively do experiments because of a missing 15x10cm rectangular white box.

Importantly, I would like to quote a great scientist and philosopher from Marseille who told me once, when this adventure was about to become a nightmare: *"When you are in the shit, you need to learn how to swim without breathing"*. Everyone is free to make up their own mind about this sentence but for me, it sums up this second year very well.

September 2021 - December 2022

Year 3: The (long) final stretch

This third and last year was certainly the most exciting and the most tiring one. And it started on the hub caps. After receiving a new StimTrack during the month of September, the experiments started again and from October to December, I did not stop doing experiments. During the month of December, I was doing almost daily round trips between the Conservatoire de Nice and the lab at Inria to collect data from student and professional musicians. It was the most tiring period of the PhD but also the most socially interesting because I took time to discuss with the participants in order to explain to them what I was doing and answer their (sometimes numerous) questions.

In parallel we started analyzing the data of the first participants. The results were promising and we decided to present them at the virtual APAN conference in November 2021. For a first poster presentation experience I have to say that it was quite weird. Presenting a PDF file behind a screen does not help receiving feedback and making connections. And thanks to Covid, I MISSED THE OPPORTUNITY TO TRAVEL TO THE US! Anyways, I received a few good feedbacks that motivated me and made me realize that our questions were relevant.

At the beginning of 2022, we concentrated our efforts on analyzing the EEG data of all the participants and we gave up on pupillometry data because we realized that it would have been too short to analyze both EEG and pupillometry data before the end of the summer. Statistical analysis was clearly not my favorite part but with some time (around 5 months), a good amount of motivation and the right people, the results were there. And they were apparently not so bad! At the beginning of June, we started to write two articles and we did a poster that I presented at the CHS-COM 2022 conference that started in Sweden, in the city of Linköping in the middle of June. It was really stimulating because I met a lot of researchers that are working on AAD. During the four days of the conference, I have been able to discuss the strength and weakness of our protocol and I received several positive feedbacks about the results. But something that I will remember for a long time is that every single place is closed after 10 pm! Even the bars! What a bizarre country...

Back in the South of France, after some weeks, we finished one of the two articles and submitted it to *The Journal of Cognitive Neuroscience*. While I am writing this paragraph, it is still under review but I have good hope that it will be published before I finish my PhD. In parallel, at the beginning of the summer, I started to work on this manuscript. If everything goes well, we should have submitted the third article before the month of October and I should have finished those lines by the end of September 2022, three years after starting one of the most difficult, stressful, tiring and exciting adventures of my life!

BIBLIOGRAPHY

- Abdel-Aziz, M., Rashed, M., Khalifa, B., Talaat, A., & Nassar, A. (2014). Eosinophilic granuloma of the temporal bone in children. *Journal of Craniofacial Surgery*, 25(3), 1076-1078.
- Adams, E. J., Nguyen, A. T., & Cowan, N. (2018). Theories of working memory: Differences in definition, degree of modularity, role of attention, and purpose. In *Language, Speech, and Hearing Services in Schools* (Vol. 49, Issue 3, pp. 340-355). https://doi.org/10.1044/2018_LSHSS-17-0114
- 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.v
- Akram, S., Presaccoc, A., Simon, J. Z., Shamma, S. A., & Babadi, B. (2016). Robust decoding of selective auditory attention from MEG in a competing-speaker environment via state-space modeling. *NeuroImage*, 124(1), 139-148.
- Akram, S., Simon, J. Z., Shamma, S., & Babadi, B. (2014). A state-space model for decoding auditory attentional modulation from MEG in a competing-speaker environment. *Advances in Neural Information Processing Systems*, 1(January), 460-468.
- Alain, C., Arnott, S. R., & Picton W., T. (2001). Bottom-Up and Top-Down Influences on Auditory Scene Analysis: Evidence From Event-Related Brain Potentials. *Journal of Experimental Psychology: Human Perception and Performance*, 27(5), 1072-1089.
- Alain, C., & Bernstein, L. J. (2015). AUDITORY SCENE ANALYSIS:TALES FROM COGNITIVE NEUROSCIENCES. *Music Perception: An Interdisciplinary Journal*, 1(33), 70-82.
- Alain, C., & McDonald, K. L. (2007). Age-related differences in neuromagnetic brain activity underlying concurrent sound perception. *Journal of Neuroscience*, 27(6), 1308-1314. <https://doi.org/10.1523/JNEUROSCI.5433-06.2007>
- Alain, C., & Woods, D. (1994). Signal clustering modulates auditory cortical activity in humans. *Perception & Psychophysics*, 56(5), 501-516. <https://doi.org/10.1145/2159365.2159412>
- Albouy, P., Benjamin, L., Morillon, B., & Zatorre, R. J. (2020). Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. *Science*, 367(6481), 1043-1047. <https://doi.org/10.1126/science.aaz3468>
- Alho, K., Töttölä, K., Reinikainen, K., Sams, M., & Näätänen, R. (1987). Brain mechanism of selective listening reflected by event-related potentials. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 68(6), 458-470.
- Alho, K., Sams, M., Paavilainen, P., & Näätänen, R. (1986a). Small pitch separation and the selective-attention effect on the ERP. *Psychophysiology*, 23(2), 189-197.
- Alho, K., Paavilainen, P., Reinikainen, K., Sams, M., & Näätänen, R. (1986b). Separability of different negative components of the event-related potential associated with auditory stimulus processing. *Psychophysiology*, 23(6), 613-623.
- Alho, K., Woods, D. L., & Algazi, A. (1994). Processing of auditory stimuli during auditory and visual attention as revealed by event-related potentials. *Psychophysiology*, 31(5), 469-479.
- Alickovic, E., Lunner, T., Gustafsson, F., & Ljung, L. (2019). A Tutorial on Auditory Attention Identification Methods. *Frontiers in Neuroscience*, 13(March), 1-17. <https://doi.org/10.3389/fnins.2019.00153>
- Alloway, T. P., Gathercole, S. E., Willis, C., & Adams, A. M. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology*, 87(2), 85-106. <https://doi.org/10.1016/j.jecp.2003.10.002>

- Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable?. *Child development*, 77(6), 1698-1716.
- Allport, A. (1987). Selection for action: Some behavioral and neurophysiological considerations of attention and action. *Perspectives on perception and action*, 15, 395-419.
- Allport, A. (1993). Attention and control: Have we been asking the wrong questions? A critical review of twenty-five years.
- Almudena, G., Santapau, M., & Gonzalez, Julian, J. (2020). EEG Analysis during Music Perception. In *Electroencephalography*.
<http://dx.doi.org/10.1039/C7RA00172J><https://www.intechopen.com/books/advanced-biometric-technologies/liveness-detection-in-biometrics><http://dx.doi.org/10.1016/j.colsurfa.2011.12.014>
- Amunts, K., Morosan, P., Hilbig, H., & Zilles, K. (2012). Auditory System. In *The Human Nervous System*.
<https://doi.org/10.1016/B978-0-12-374236-0.10036-7>
- Anderson, B. (2011). There is no such thing as attention. *Frontiers in Psychology*, 2(SEP), 1-8.
<https://doi.org/10.3389/fpsyg.2011.00246>
- Anderson, B. (2021). Stop paying attention to "attention." *Wiley Interdisciplinary Reviews: Cognitive Science*, March, 1-6. <https://doi.org/10.1002/wcs.1574>
- Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Bonelli, S., Golfetti, A., Pozzi, S., Imbert, J. P., Granger, G., Benhacene, R., & Babiloni, F. (2016). Adaptive automation triggered by EEG-based mental workload index: A passive brain-computer interface application in realistic air traffic control environment. *Frontiers in Human Neuroscience*, 10(OCT2016), 1-13. <https://doi.org/10.3389/fnhum.2016.00539>
- Arnott, S. R., Binns, M. A., Grady, C. L., & Alain, C. (2004). Assessing the auditory dual-pathway model in humans. *NeuroImage*, 22(1), 401-408. <https://doi.org/10.1016/j.neuroimage.2004.01.014>
- Arslan, B., Brause, A., Castet, J., Léhembre, R., Simon, C., Filatriau, J. J., & Noirhomme, Q. (2006). A real time music synthesis environment driven with biological signals. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2, 1172-1175.
<https://doi.org/10.1109/icassp.2006.1660557>
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *Psychology of learning and motivation (Vol. 2, pp. 89-195)*. Academic Press.
- Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, 5(3), 119-126. [https://doi.org/10.1016/S1364-6613\(00\)01593-X](https://doi.org/10.1016/S1364-6613(00)01593-X)
- Baddeley, A. (1996). The fractionation of working memory. *Proceedings of the National Academy of Sciences of the United States of America*, 93(24), 13468-13472.
<https://doi.org/10.1073/pnas.93.24.13468>
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation (Vol. 8, pp. 47-89)*. Academic press.
- Baddeley, A. (2000). The episodic buffer: a new component of working memory?. *Trends in cognitive sciences*, 4(11), 417-423.
- Bailey, C. E. (2007). Cognitive accuracy and intelligent executive function in the brain and in business. *Annals of the New York Academy of Sciences*, 1118(1), 122-141.
- Baird, B., Smallwood, J., & Schooler, J. W. (2010). I can shake that feeling: Positive mind-wandering prevents the deterioration of mood. In Poster presented at Towards a Science of Consciousness conference, Tucson, AZ.
- Baird, B., Smallwood, J., Mrazek, M. D., Kam, J. W. Y., Franklin, M. S., & Schooler, J. W. (2012). Inspired by Distraction: Mind Wandering Facilitates Creative Incubation. *Psychological Science*, 23(10), 1117-1122.
<https://doi.org/10.1177/0956797612446024>

- Baltzell, L. S., Srinivasan, R., & Richards, V. M. (2017). The effect of prior knowledge and intelligibility on the cortical entrainment response to speech. *Journal of Neurophysiology*, 118(6), 3144–3151. <https://doi.org/10.1152/jn.00023.2017>
- Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., Heinze, H. J., & Altenmüller, E. (2006). Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. *NeuroImage*, 30(3), 917–926. <https://doi.org/10.1016/j.neuroimage.2005.10.044>
- Barker, B. A., & Newman, R. S. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition*, 94(2), 45–53. <https://doi.org/10.1016/j.cognition.2004.06.001>
- Barron, E., Riby, L. M., Greer, J., & Smallwood, J. (2011). Absorbed in thought: The effect of mind wandering on the processing of relevant and irrelevant events. *Psychological Science*, 22(5), 596–601. <https://doi.org/10.1177/0956797611404083>
- Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time Constraints and Resource Sharing in Adults' Working Memory Spans. *Journal of Experimental Psychology: General*, 133(1), 83–100. <https://doi.org/10.1037/0096-3445.133.1.83>
- Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and Cognitive Load in Working Memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, 33(3), 570–585. <https://doi.org/10.1037/0278-7393.33.3.570>
- Barrouillet, P., Gavens, N., Vergauwe, E., Gaillard, V., & Camos, V. (2009). Working Memory Span Development: A Time-Based Resource-Sharing Model Account. *Developmental Psychology*, 45(2), 477–490. <https://doi.org/10.1037/a0014615>
- Battistoni, E., Stein, T., & Peelen, M. V. (2017). Preparatory attention in visual cortex. *Annals of the New York Academy of Sciences*, 1396(1), 92–107. <https://doi.org/10.1111/nyas.13320>
- Baumann, S., Koeneke, S., Schmidt, C. F., Meyer, M., Lutz, K., & Jancke, L. (2007). A network for audio-motor coordination in skilled pianists and non-musicians. *Brain Research*, 1161(1), 65–78. <https://doi.org/10.1016/j.brainres.2007.05.045>
- Baumann, S., Meyer, M., & Jäncke, L. (2008). Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. *Journal of Cognitive Neuroscience*, 20(12), 2238–2249. <https://doi.org/10.1162/jocn.2008.20157>
- Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The Complexities of Complex Span: Explaining Individual Differences in Working Memory in Children and Adults. *Journal of Experimental Psychology: General*, 132(1), 71–92. <https://doi.org/10.1037/0096-3445.132.1.71>
- Belin, P., & Zatorre, R. J. (2000). 'What', 'where' and 'how' in auditory cortex. *Nature neuroscience*, 3(10), 965–966.
- Belo, J., Clerc, M., & Schön, D. (2021). EEG-Based Auditory Attention Detection and Its Possible Future Applications for Passive BCI. *Frontiers in Computer Science*, 3(April), 1–9. <https://doi.org/10.3389/fcomp.2021.661178>
- Belo, J., Clerc, M., & Schön, D. (2022). Attentional inhibition ability predicts neural representation during challenging auditory streaming. *BioRxiv Preprint*, 1–29.
- Bender, W. N. (2004). *Learning disabilities: Characteristics, identification, and teaching strategies*. Allyn & Bacon.
- Bendixen, A. (2014). Predictability effects in auditory scene analysis: A review. *Frontiers in Neuroscience*, 8(MAR), 1–16. <https://doi.org/10.3389/fnins.2014.00060>
- Bendixen, A., Bohm, T. M., Szalárdy, O., Mill, R., Denham, S. L., & Winkler, I. (2013). Different roles of similarity and predictability in auditory stream segregation. *Learning and Perception*, 5, 37–54. <https://doi.org/10.1556/LP.5.2013.Suppl2.4>

- Bendixen, A., Denham, S. L., Gyimesi, K., & Winkler, I. (2010). Regular patterns stabilize auditory streams. *The Journal of the Acoustical Society of America*, 128(6), 3658–3666. <https://doi.org/10.1121/1.3500695>
- Bengtsson, S. L., Nagy, Z., Skare, S., Forsman, L., Forssberg, H., & Ullén, F. (2005). Extensive piano practicing has regionally specific effects on white matter development. *Nature Neuroscience*, 8(9), 1148–1150. <https://doi.org/10.1038/nn1516>
- Benz, S., Sellaro, R., Hommel, B., & Colzato, L. S. (2016). Music makes the world go round: The impact of musical training on non-musical cognitive functions—a review. *Frontiers in Psychology*, 6(JAN), 1–5. <https://doi.org/10.3389/fpsyg.2015.02023>
- Bergman Nutley, S., Darki, F., & Klingberg, T. (2014). Music practice is associated with development of working memory during childhood and adolescence. *Frontiers in Human Neuroscience*, 7(JAN), 1–9. <https://doi.org/10.3389/fnhum.2013.00926>
- Bermudez, P., Lerch, J. P., Evans, A. C., & Zatorre, R. J. (2009). Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cerebral Cortex*, 19(7), 1583–1596. <https://doi.org/10.1093/cercor/bhn196>
- Berti, S., & Schröger, E. (2003). Working memory controls involuntary attention switching: Evidence from an auditory distraction paradigm. *European Journal of Neuroscience*, 17(5), 1119–1122. <https://doi.org/10.1046/j.1460-9568.2003.02527.x>
- 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. *Trends in Amplification*, 17(2), 75–93. <https://doi.org/10.1177/1084713813495459>
- Besson, M., & Schön, D. (2012). Comparison Between Language and Music. *The Cognitive Neuroscience of Music*, 232–258. <https://doi.org/10.1093/acprof:oso/9780198525202.003.0018>
- Best, V., Gallun, F. J., Mason, C. R., Jr, G. K., & Barabra, G. (2010). The impact of noise and hearing loss on the processing of simultaneous sentences. *Ear and Hearing*, 31(2), 213–220. <https://doi.org/10.1097/AUD.0b013e3181c34ba6>
- Best, V., Ozmeral, E. J., Kopčo, N., & Shinn-Cunningham, B. G. (2008). Object continuity enhances selective auditory attention. *Proceedings of the National Academy of Sciences of the United States of America*, 105(35), 13174–13178. <https://doi.org/10.1073/pnas.0803718105>
- Bey, C., & Adams, S. M. C. (2002). Schema-Based Auditory Scene Analysis. *Perception & Psychophysics*, 64(5), 844–854.
- Bialystok, E., & DePape, A. M. (2009). Musical Expertise, Bilingualism, and Executive Functioning. *Journal of Experimental Psychology: Human Perception and Performance*, 35(2), 565–574. <https://doi.org/10.1037/a0012735>
- Bianchi, F., Santurette, S., Wendt, D., & Dau, T. (2016). Pitch Discrimination in Musicians and Non-Musicians: Effects of Harmonic Resolvability and Processing Effort. *JARO - Journal of the Association for Research in Otolaryngology*, 17(1), 69–79. <https://doi.org/10.1007/s10162-015-0548-2>
- Bidelman, G. M., & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Research*, 1355, 112–125. <https://doi.org/10.1016/j.brainres.2010.07.100>
- Bidelman, G. M., & Yoo, J. (2020). Musicians Show Improved Speech Segregation in Competitive, Multi-Talker Cocktail Party Scenarios. *Frontiers in Psychology*, 11(August), 1–11. <https://doi.org/10.3389/fpsyg.2020.01927>
- Bidet-Caulet, A., Fischer, C., Besle, J., Aguera, P. E., Giard, M. H., & Bertrand, O. (2007). Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. *Journal of Neuroscience*, 27(35), 9252–9261. <https://doi.org/10.1523/JNEUROSCI.1402-07.2007>

- Bidet-Caulet, A., Mikyska, C., & Knight, R. T. (2010). Load effects in auditory selective attention: Evidence for distinct facilitation and inhibition mechanisms. *NeuroImage*, 50(1), 277–284. <https://doi.org/10.1016/j.neuroimage.2009.12.039>
- Biesmans, W., Vanthornhout, J., Wouters, J., Moonen, M., Francart, T., & Bertrand, A. (2015). Comparison of speech envelope extraction methods for EEG-based auditory attention detection in a cocktail party scenario. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2015-Novem*, 5155–5158. <https://doi.org/10.1109/EMBC.2015.7319552>
- Bigand, E., McAdams, S., & Forêt, S. (2000). Divided attention in music. *International Journal of Psychology*, 35(5–6), 270–278. <https://doi.org/10.1080/002075900750047987>
- Bigand, E., & Poulin-Charronnat, B. (2006). Are we “experienced listeners”? A review of the musical capacities that do not depend on formal musical training. *Cognition*, 100(1), 100–130. <https://doi.org/10.1016/j.cognition.2005.11.007>
- Billig, A. J., & Carlyon, R. P. (2016). Automaticity and primacy of auditory streaming: Concurrent subjective and objective measures. *Journal of Experimental Psychology: Human Perception and Performance*, 42(3), 339–353. <https://doi.org/10.1037/xhp0000146>
- Blain, S., Talamini, F., Fornoni, L., Bidet-Caulet, A., & Caclin, A. (2022). Shared cognitive resources between memory and attention during sound sequence encoding. *Attention, Perception & Psychophysics*, 84(2), 739–759.
- Blank, H., Wieland, N., & von Kriegstein, K. (2014). Person recognition and the brain: Merging evidence from patients and healthy individuals. *Neuroscience and Biobehavioral Reviews*, 47, 717–734. <https://doi.org/10.1016/j.neubiorev.2014.10.022>
- Bledowski, C., Kaiser, J., & Rahm, B. (2010). Basic operations in working memory: Contributions from functional imaging studies. *Behavioural Brain Research*, 214(2), 172–179. <https://doi.org/10.1016/j.bbr.2010.05.041>
- Bleichner, M. G., Mirkovic, B., & Debener, S. (2016). Identifying auditory attention with ear-EEG: CEEGrid versus high-density cap-EEG comparison. *Journal of Neural Engineering*, 13(6). <https://doi.org/10.1088/1741-2560/13/6/066004>
- Boebinger, D., Evans, S., Rosen, S., Lima, C. F., Manly, T., & Scott, S. K. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *The Journal of the Acoustical Society of America*, 137(1), 378–387. <https://doi.org/10.1121/1.4904537>
- Boehnke, S. E., & Phillips, D. P. (2005). The relation between auditory temporal interval processing and sequential stream segregation examined with stimulus laterality differences. *Perception and Psychophysics*, 67(6), 1088–1101. <https://doi.org/10.3758/BF03193634>
- Bonin, T., & Smilek, D. (2016). Inharmonic music elicits more negative affect and interferes more with a concurrent cognitive task than does harmonic music. *Attention, Perception, and Psychophysics*, 78(3), 946–959. <https://doi.org/10.3758/s13414-015-1042-y>
- Borella, E., Carretti, B., & Pelegrina, S. (2010). The Specific Role of Inhibition in Reading Comprehension in Good and Poor Comprehenders. *Journal of Learning Disabilities*, 43(6), 541–552. <https://doi.org/10.1177/0022219410371676>
- Botvinick, M. M., Braver, T. S., Barch, D., Carter, C. S., & Cohen, J. D. (2001). Conflict Monitoring and Cognitive Control. *Psychological Review*, 108(3), 624–652. <https://doi.org/10.11646/zootaxa.4127.2.2>
- Boucher, L., Palmeri, T. J., Logan, G. D., & Schall, J. D. (2007). Inhibitory control in mind and brain: An interactive race model of countermanding saccades. *Psychological Review*, 114(2), 376–397. <https://doi.org/10.1037/0033-295X.114.2.376>
- Bregman, A. S. (1990). Auditory Scene Analysis: The Perceptual Organization of Sound. *The Journal of the Acoustical Society of America*, 95(2), 1177–1178. <https://doi.org/10.1121/1.408434>

- Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press.
- Bregman, A. S., & Levitan, R. (1983). Stream segregation based on fundamental frequency and spectral peak. 1: Effects of shaping by filters. Unpublished manuscript, Psychology Department, McGill University.
- Bregman, A. S., Ahad, P. A., & Van Loon, C. (2001). Stream segregation of narrow-band noise bursts. *Perception and Psychophysics*, 63(5), 790–797. <https://doi.org/10.3758/BF03194438>
- Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, 89(2), 244–249. <https://doi.org/10.1037/h0031163>
- Broadbent, D. E. (1952). Listening to one of two synchronous messages. *Journal of experimental psychology*, 44(1), 51.
- Broadbent, D. E. (1962). Attention and the perception of speech. *Scientific American*, 206(4), 143-153.
- Broadbent, D. E. (1958). *Perception and communication*. New York: Pergamon Press.
- Brungart, D. S. (2001). 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>
- Bugos, J. A., Perlstein, W. M., McCrae, C. S., Brophy, T. S., & Bedenbaugh, P. H. (2007). Individualized Piano Instruction enhances executive functioning and working memory in older adults. *Aging and Mental Health*, 11(4), 464–471. <https://doi.org/10.1080/13607860601086504>
- Bulger, E., Shinn-Cunningham, B. G., & Noyce, A. L. (2021). Distractor probabilities modulate flanker task performance. *Attention, Perception, and Psychophysics*, 83(2), 866–881. <https://doi.org/10.3758/s13414-020-02151-7>
- Burgess, P. W., & Simons, J. S. (2015). Theories of frontal lobe executive function: Clinical applications. In *Effectiveness of Rehabilitation for Cognitive Deficits* (pp. 211–232). <https://doi.org/10.1093/acprof:oso/9780198526544.003.0018>
- Burnham, B. R., Sabia, M., & Langan, C. (2014). Components of working memory and visual selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, 40(1), 391–403. <https://doi.org/10.1037/a0033753>
- Burunat, I., Alluri, V., Toiviainen, P., Numminen, J., & Brattico, E. (2014). Dynamics of brain activity underlying working memory for music in a naturalistic condition. *Cortex*, 57(M), 254–269. <https://doi.org/10.1016/j.cortex.2014.04.012>
- Buyens, W., van Dijk, B., Moonen, M., & Wouters, J. (2014). Music mixing preferences of cochlear implant recipients: A pilot study. *International journal of audiology*, 53(5), 294-301.
- Caldwell, A., & Nittrouer, S. (2013). Speech Perception in Noise by Children With Cochlear Implants. *Journal of Speech, Language, and Hearing Research*, 54(1), 1–33. [https://doi.org/10.1044/1092-4388\(2012/11-0338\).Speech](https://doi.org/10.1044/1092-4388(2012/11-0338).Speech)
- Cammoun, L., Thiran, J. P., Griffa, A., Meuli, R., Hagmann, P., & Clarke, S. (2015). Intrahemispheric cortico-cortical connections of the human auditory cortex. *Brain Structure and Function*, 220(6), 3537–3553. <https://doi.org/10.1007/s00429-014-0872-z>
- Campbell, K. L., Lustig, C., & Hasher, L. (2020). Aging and inhibition: Introduction to the special issue. *Psychology and Aging*, 35(5), 605–613. <https://doi.org/10.1037/pag0000564>
- Cant, N. B., & Hyson, R. L. (1992). Projections from the lateral nucleus of the trapezoid body to the medial superior olivary nucleus in the gerbil. *Hearing research*, 58(1), 26-34.

- Cantisani, G., Essid, S., & Richard, G. (2019). EEG-Based decoding of auditory attention to a target instrument in polyphonic music. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2019-October, 80–84. <https://doi.org/10.1109/WASPAA.2019.8937219>
- Carlyon, R. P. (2004). How the brain separates sounds. *Trends in Cognitive Sciences*, 8(10), 465–471. <https://doi.org/10.1016/j.tics.2004.08.008>
- Carlyon, R. P., Cusack, R., Foxtton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1), 115–127. <https://doi.org/10.1037/0096-1523.27.1.115>
- Carroll YI, Eichwald J, S. F. (2017). Morbidity and Mortality Weekly Report Vital Signs: Noise-Induced Hearing Loss Among Adults — United States. *Morbidity and Mortality Weekly Report*, 10(5), 139–144.
- Caspary, D. M. (1991). Superior olivary complex-Functional neuropharmacology of the principal cell types. *Neurobiology of hearing: the central auditory system*, 141.
- Cervantes Constantino, F., & Simon, J. Z. (2018). Restoration and efficiency of the neural processing of continuous speech are promoted by prior knowledge. *Frontiers in Systems Neuroscience*, 12(October), 1–11. <https://doi.org/10.3389/fnsys.2018.00056>
- Chait, M., de Cheveigné, A., Poeppel, D., & Simon, J. Z. (2010). Neural dynamics of attending and ignoring in human auditory cortex. *Neuropsychologia*, 48(11), 3262–3271. <https://doi.org/10.1016/j.neuropsychologia.2010.07.007>
- Chelazzi, L., Bisley, J. W., & Bartolomeo, P. (2018). The unconscious guidance of attention. *Cortex*, 102, 1–5. <https://doi.org/10.1016/j.cortex.2018.02.002>
- Chen, Z., & Cowan, N. (2009). Core verbal working-memory capacity: The limit in words retained without covert articulation. *Quarterly Journal of Experimental Psychology*, 62(7), 1420–1429. <https://doi.org/10.1080/17470210802453977>
- Chen, Z., & Cowan, N. (2005). Chunk Limits and Length Limits in Immediate Recall: A Reconciliation. *Journal of Experimental Psychology: Learning Memory and Cognition*, 31(6), 1235–1249. <https://doi.org/10.1037/0278-7393.31.6.1235.Chunk>
- 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.
- Cho, B. H., Lee, J. M., Ku, J. H., Jang, D. P., Kim, J. S., Kim, I. Y., Lee, J. H., & Kim, S. I. (2002). Attention enhancement system using virtual reality and EEG biofeedback. *Proceedings - Virtual Reality Annual International Symposium*, July 2014, 156–163. <https://doi.org/10.1109/vr.2002.996518>
- Christakis, D. A., Zimmerman, F. J., DiGiuseppe, D. L., & McCarty, C. A. (2004). Early Television Exposure and Subsequent Attentional Problems in Children. *Pediatrics*, 113(4 I), 708–713. <https://doi.org/10.1542/peds.113.4.708>
- Christiansen, S. K., Jepsen, M. L., & Dau, T. (2014). Effects of tonotopicity, adaptation, modulation tuning, and temporal coherence in “primitive” auditory stream segregation. *The Journal of the Acoustical Society of America*, 135(1), 323–333.
- Christison-Lagay, K. L., Gifford, A. M., & Cohen, Y. E. (2015). Neural correlates of auditory scene analysis and perception. *International Journal of Psychophysiology*, 95(2), 238–245. <https://doi.org/10.1016/j.ijpsycho.2014.03.004>
- Christoff, K., Mills, C., Andrews-Hanna, J. R., Irving, Z. C., Thompson, E., Fox, K. C. R., & Kam, J. W. Y. (2018). Mind-Wandering as a Scientific Concept: Cutting through the Definitional Haze. *Trends in Cognitive Sciences*, 22(11), 957–959. <https://doi.org/10.1016/j.tics.2018.07.004>
- Ciccarelli, G., Nolan, M., Perricone, J., Calamia, P. T., Haro, S., O’Sullivan, J., Mesgarani, N., Quatieri, T. F., & Smalt, C. J. (2019). Comparison of Two-Talker Attention Decoding from EEG with Nonlinear Neural

- Networks and Linear Methods. *Scientific Reports*, 9(1), 1–10. <https://doi.org/10.1038/s41598-019-47795-0>
- Ciocca, V. (2008). The auditory organization of complex sounds. *Frontiers in Bioscience*, 13, 148–169.
- Clayton, K. K., Swaminathan, J., Yazdanbakhsh, A., Zuk, J., Patel, A. D., & Kidd, G. (2016). Executive function, visual attention and the cocktail party problem in musicians and non-musicians. *PLoS ONE*, 11(7), 1–17. <https://doi.org/10.1371/journal.pone.0157638>
- Clayton, M. S., Yeung, N., & Cohen Kadosh, R. (2015). The roles of cortical oscillations in sustained attention. *Trends in Cognitive Sciences*, 19(4), 188–195. <https://doi.org/10.1016/j.tics.2015.02.004>
- Cocchini, G., Logie, R. H., Della Sala, S., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory and Cognition*, 30(7), 1086–1095. <https://doi.org/10.3758/BF03194326>
- Coffey, E. B. J., Mogilever, N. B., & Zatorre, R. J. (2017). Speech-in-noise perception in musicians: A review. *Hearing Research*, 352, 49–69. <https://doi.org/10.1016/j.heares.2017.02.006>
- Coleman, A., & Cervin, A. (2019). Probiotics in the treatment of otitis media. The past, the present and the future. *International Journal of Pediatric Otorhinolaryngology*, 116(August 2018), 135–140. <https://doi.org/10.1016/j.ijporl.2018.10.023>
- Coles, M. G. H., & Donchin, E. (1988). Is the P300 component a manifestation of context updating? In *Behavioral and Brain Sciences* (Vol. 11, pp. 357–427). https://www.researchgate.net/profile/Emanuel_Donchin/publication/232559111_Is_the_P300_component_a_manifestation_of_context_updating/links/55ef861108aef559dc44d610/Is-the-P300-component-a-manifestation-of-context-updating.pdf
- Colflesh, G. J. H., & Conway, A. R. A. (2007). Individual differences in working memory capacity and divided attention in dichotic listening. *Psychonomic Bulletin and Review*, 14(4), 699–703. <https://doi.org/10.3758/BF03196824>
- Connine, C. M., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word Familiarity and Frequency in Visual and Auditory Word Recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(6), 1084–1096. <https://doi.org/10.1037/0278-7393.16.6.1084>
- Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin and Review*, 8(2), 331–335. <https://doi.org/10.3758/BF03196169>
- Cowan, N. (1999). An embedded-processes model of working memory.
- Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin and Review*, 24(4), 1158–1170. <https://doi.org/10.3758/s13423-016-1191-6>
- Cowan, N. (1988). Evolving Conceptions of Memory Storage, Selective Attention, and Their Mutual Constraints Within the Human Information-Processing System. *Psychological Bulletin*, 104(2), 163–191.
- Cowan, N., Chen, Z., & Rouder, J. N. (2004). Constant capacity in an immediate serial-recall task: A logical sequel to Miller (1956). *Psychological Science*, 15(9), 634–640. <https://doi.org/10.1111/j.0956-7976.2004.00732.x>
- Cowan, N., Elliott, E. M., Saults, S. J., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, 51(1), 42–100. <https://doi.org/10.1016/j.cogpsych.2004.12.001>
- Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education*, 3(2), 63–68. <https://doi.org/10.1016/j.tine.2013.12.001>

- Crandell, C. C. (1991). Individual differences in speech recognition ability: Implications for hearing aid selection. *Ear and Hearing*, 12(6), 100s-108s.
- Crosse, M. J., Zuk, N. J., Liberto, G. M. Di, Nidiffer, A. R., Molholm, S., & Lalor, E. C. (2021). Linear Modeling of Neurophysiological Responses to Naturalistic Stimuli : Methodological Considerations for Applied Research. *PsyArXiv*.
- Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The multivariate temporal response function (mTRF) toolbox: A MATLAB toolbox for relating neural signals to continuous stimuli. *Frontiers in Human Neuroscience*, 10(NOV2016), 1–14. <https://doi.org/10.3389/fnhum.2016.00604>
- Cusack, R., Deeks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology: Human Perception and Performance*, 30(4), 643–656. <https://doi.org/10.1037/0096-1523.30.4.643>
- Cusack, R., & Roberts, B. (1999). Effects of similarity in bandwidth on the auditory sequential streaming of two-tone complexes. *Perception*, 28(10), 1281–1289. <https://doi.org/10.1068/p2804>
- D'Souza, A. A., Moradzadeh, L., & Wiseheart, M. (2018). Musical training, bilingualism, and executive function: working memory and inhibitory control. *Cognitive Research: Principles and Implications*, 3(1). <https://doi.org/10.1186/s41235-018-0095-6>
- da Costa, S., van der Zwaag, W., Marques, J. P., Frackowiak, R. S. J., Clarke, S., & Saenz, M. (2011). Human primary auditory cortex follows the shape of Heschl's Gyrus. *Journal of Neuroscience*, 31(40), 14067–14075. <https://doi.org/10.1523/JNEUROSCI.2000-11.2011>
- Da Silva, F. H. L. (2010). EEG : Origin and Measurement. *EEG - FMRI: Physiological Basis, Technique, and Applications*, 19–37. <https://doi.org/10.1007/978-3-540-87919-0>
- Dalton, P., & Hughes, R. W. (2014). Auditory attentional capture: Implicit and explicit approaches. *Psychological Research*, 78(3), 313–320. <https://doi.org/10.1007/s00426-014-0557-5>
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior*, 19(4), 450-466.
- Daniel, K., B.Alan, K., David, S. A., Norbert, S., & Arthur, S. A. (2004). A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method. *Science*, 306(5702), 1776–1780.
- Danielsen, A., Nymoen, K., Langerød, M. T., Jacobsen, E., Johansson, M., & London, J. (2022). Sounds familiar(?): Expertise with specific musical genres modulates timing perception and micro-level synchronization to auditory stimuli. *Attention, Perception, and Psychophysics*, 84(2), 599–615. <https://doi.org/10.3758/s13414-021-02393-z>
- Danielsen, A., Nymoen, K., Anderson, E., Câmara, G. S., Langerød, M. T., Thompson, M. R., & London, J. (2019). Where is the beat in that note? Effects of attack, duration, and frequency on the perceived timing of musical and quasi-musical sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 45(3), 402.
- Danneels, M., Van Hecke, R., Leyssens, L., Degeest, S., Cambier, D., Van De Berg, R., Van Rompaey, V., & Maes, L. (2020). 2BALANCE: A cognitive-motor dual-task protocol for individuals with vestibular dysfunction. *BMJ Open*, 10(7), 1–11. <https://doi.org/10.1136/bmjopen-2020-037138>
- Dannenbring, G. L., & Bregman, A. S. (1978). Streaming vs. fusion of sinusoidal components of complex tones. *Perception & Psychophysics*, 24(4), 369–376. <https://doi.org/10.3758/BF03204255>
- Darwin, C. J. (1984). Perceiving vowels in the presence of another sound: Constraints on formant perception. *Journal of the Acoustical Society of America*, 76(6), 1636–1647. <https://doi.org/10.1121/1.391610>

- Darwin, C. J. (2006). Contributions of binaural information to the separation of different sound sources: Contribuciones de la información binaural en la separación de diferentes fuentes sonoras. *International journal of audiology*, 45(sup1), 20-24.
- Darwin, C. J. (2008). Listening to speech in the presence of other sounds. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), 1011–1021. <https://doi.org/10.1098/rstb.2007.2156>
- Darwin, C. J., & Hukin, R. W. (1999). Auditory objects of attention: The role of interaural time differences. *Journal of Experimental Psychology: Human Perception and Performance*, 25(3), 617–629. <https://doi.org/10.1037/0096-1523.25.3.617>
- Darwin, C. J., & Hukin, R. W. (1998). Perceptual segregation of a harmonic from a vowel by interaural time difference in conjunction with mistuning and onset asynchrony. *The Journal of the Acoustical Society of America*, 103(2), 1080–1084. <https://doi.org/10.1121/1.421221>
- de Cheveigné, A., Wong, D. E., Di Liberto, G. M., Hjortkjær, J., Slaney, M., & Lalor, E. (2018). Decoding the auditory brain with canonical component analysis. *NeuroImage*, 172(September 2017), 206–216. <https://doi.org/10.1016/j.neuroimage.2018.01.033>
- De Fockert, J. W., Rees, G., Frith, C. D., & Lavie, N. (2001). The role of working memory in visual selective attention. *Science*, 291(5509), 1803–1806. <https://doi.org/10.1126/science.1056496>
- De Santis, L., Clarke, S., & Murray, M. M. (2007). Automatic and intrinsic auditory “what” and “where” processing in humans revealed by electrical neuroimaging. *Cerebral Cortex*, 17(1), 9–17. <https://doi.org/10.1093/cercor/bhj119>
- de Taillez, T., Kollmeier, B., & Meyer, B. T. (2017). Machine learning for decoding listeners’ attention from EEG evoked by continuous speech. *European Journal of Neuroscience*, 51, 1234–1241. <https://doi.org/10.1111/ijlh.12426>
- Deckers, L., Das, N., Ansari, A. H., Bertrand, A., & Francart, T. (2018). EEG-based detection of the attended speaker and the locus of auditory attention with convolutional neural networks. *BioRxiv*, 475673. <https://doi.org/10.1101/475673>
- Deike, S., Gaschler-Markefski, B., Brechmann, A., & Scheich, H. (2004). Auditory stream segregation relying on timbre involves left auditory cortex. *NeuroReport*, 15(9), 1511–1514. <https://doi.org/10.1097/01.wnr.0000132919.12990.34>
- Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Psychology*, 18, 193–222. <https://doi.org/10.1146/annurev-psych-122414-033400>
- Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, 70(1), 51–60. <https://doi.org/10.1037/h0042712>
- Devergie, A., Grimault, N., Tillmann, B., & Berthommier, F. (2010). Effect of rhythmic attention on the segregation of interleaved melodies. *The Journal of the Acoustical Society of America*, 128(1), EL1–EL7. <https://doi.org/10.1121/1.3436498>
- Di Liberto, G. M., Marion, G., & Shamma, S. A. (2021). Accurate Decoding of Imagined and Heard Melodies. *Frontiers in Neuroscience*, 15(August), 1–11. <https://doi.org/10.3389/fnins.2021.673401>
- Di Liberto, G. M., Pelofi, C., Bianco, R., Patel, P., Mehta, A. D., Herrero, J. L., de Cheveigné, A., Shamma, S., & Mesgarani, N. (2020a). Cortical encoding of melodic expectations in human temporal cortex. *ELife*, 9, 1–26. <https://doi.org/10.7554/eLife.51784>
- Di Liberto, G. M., Pelofi, C., Shamma, S., & de Cheveigné, A. (2020b). Musical expertise enhances the cortical tracking of the acoustic envelope during naturalistic music listening. *Acoustical Science and Technology*, 41(1), 361–364. <https://doi.org/10.1250/ast.41.361>
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135–168. <https://doi.org/10.1146/annurev-psych-113011-143750>

- Diamond, A. (2020). Executive functions. In *Handbook of clinical neurology* (Vol. 173, pp. 225-240). Elsevier.
- Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences of the United States of America*, 109(29), 11854–11859. <https://doi.org/10.1073/pnas.1205381109>
- Ding, Y., Gray, K., Forrence, A., Wang, X., & Huang, J. (2018). A behavioral study on tonal working memory in musicians and non-musicians. *PLoS ONE*, 13(8), 6–8. <https://doi.org/10.1371/journal.pone.0201765>
- Disbergen, N. R., Valente, G., Formisano, E., & Zatorre, R. J. (2018). Assessing top-down and bottom-up contributions to auditory stream segregation and integration with polyphonic music. *Frontiers in Neuroscience*, 12(MAR). <https://doi.org/10.3389/fnins.2018.00121>
- Dixon, P., & Bortolussi, M. (2013). Construction, integration, and mind wandering in readings. *Canadian Journal of Experimental Psychology*, 67(1), 1–10. <https://doi.org/10.1037/a0031234>
- Doelling, K. B., & Poeppel, D. (2015). Cortical entrainment to music and its modulation by expertise. *Proceedings of the National Academy of Sciences of the United States of America*, 112(45), E6233–E6242. <https://doi.org/10.1073/pnas.1508431112>
- Donald, M. W. (1987). The timing and polarity of different attention-related ERP changes inside and outside of the attentional focus. *Electroencephalography and clinical neurophysiology. Supplement*, 40, 81-86.
- Donohue, S. E., Liotti, M., Perez, R., & Woldorff, M. G. (2012). Is conflict monitoring supramodal? Spatiotemporal dynamics of cognitive control processes in an auditory Stroop task. *Cognitive, Affective and Behavioral Neuroscience*, 12(1), 1–15. <https://doi.org/10.3758/s13415-011-0060-z>
- Dowling, W. J. (1973). The perception of interleaved melodies. *Cognitive Psychology*, 5(3), 322–337. [https://doi.org/10.1016/0010-0285\(73\)90040-6](https://doi.org/10.1016/0010-0285(73)90040-6)
- Drewe, E. A. (1975). Go - No Go Learning After Frontal Lobe Lesions in Humans. *Cortex*, 11(1), 8–16. [https://doi.org/10.1016/S0010-9452\(75\)80015-3](https://doi.org/10.1016/S0010-9452(75)80015-3)
- Dryden, A., Allen, H. A., Henshaw, H., & Heinrich, A. (2017). The Association Between Cognitive Performance and Speech-in-Noise Perception for Adult Listeners: A Systematic Literature Review and Meta-Analysis. *Trends in Hearing*, 21, 1–21. <https://doi.org/10.1177/2331216517744675>
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Hustonb, A. C., Klebanov, P., Pagani, L., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School Readiness and Later Achievement. *Developmental Psychology*, 43(6), 1428.
- Dupoux, E., Gardelle, V. de, & Kouider, S. (2008). Subliminal speech perception and auditory streaming. *Cognition*, 109(2), 267–273. <https://doi.org/10.1016/j.cognition.2008.06.012>
- E, A., L, A.-V., & SA, H. (2000). The role of spatial selective attention in working memory for locations: evidence from event-related potentials. *Journal of Cognitive Neuroscience*, 12(5), 840–847.
- Edkins, G. D., & Pollock, C. M. (1997). The influence of sustained attention on railway accidents. *Accident Analysis & Prevention*, 29(4), 533-539.
- Egner, T., Delano, M., & Hirsch, J. (2007). Separate conflict-specific cognitive control mechanisms in the human brain. *NeuroImage*, 35(2), 940–948. <https://doi.org/10.1016/j.neuroimage.2006.11.061>
- Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience*, 8(12), 1784–1790. <https://doi.org/10.1038/nn1594>
- Elhilali, M., Ma, L., Micheyl, C., Oxenham, A. J., & Shamma, S. A. (2009). Temporal Coherence in the Perceptual Organization and Cortical Representation of Auditory Scenes. *Neuron*, 61(2), 317–329. <https://doi.org/10.1016/j.neuron.2008.12.005>

- Elhilali, M., Xiang, J., Shamma, S. A., & Simon, J. Z. (2009). Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. *PLoS Biology*, 7(6). <https://doi.org/10.1371/journal.pbio.1000129>
- Elliot, C. D., Smith, P., & McCulloch, K. (1997). British Abilities Scale second edition (BAS II) technical manual. Windsor, Berkshire: NFER-Nelson.
- Enger, T. (2007). Congruency sequence effects. *Cognitive, Affective, & Behavioral Neuroscience*, 7(4), 380–390.
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current directions in psychological science*, 11(1), 19-23.
- Engle, R. W., & Kane, M. J. (2004). Executive Attention, Working Memory Capacity, and a Two-Factor Theory of Cognitive Control. *Psychology of Learning and Motivation - Advances in Research and Theory*, 44, 145–199. [https://doi.org/10.1016/S0079-7421\(03\)44005-X](https://doi.org/10.1016/S0079-7421(03)44005-X)
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological review*, 102(2), 211.
- Eriksen, C. W. (1995). The flankers task and response competition: A useful tool for investigating a variety of cognitive problems. *Visual Cognition*, 2(2-3), 101-118.
- Eriksen, B., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, 16(1), 143–149. https://doi.org/10.1007/978-3-319-57111-9_9085
- Eriksen, C. W., & St James, J. D. S. (1986). Visual attention within and around the field of focal attention. *Perception & Psychophysics*, 40(4), 225–240.
- Escobar, J., Mussoi, B. S., & Silberer, A. B. (2020). The Effect of Musical Training and Working Memory in Adverse Listening Situations. *Ear and Hearing*, 278–288. <https://doi.org/10.1097/AUD.0000000000000754>
- Espy, K. A. (2004). Using developmental, cognitive, and neuroscience approaches to understand executive control in young children. *Developmental Neuropsychology*, 26(1), 379–384. https://doi.org/10.1207/s15326942dn2601_1
- Esterman, M., & Rothlein, D. (2019). Models of sustained attention. *Current Opinion in Psychology*, 29, 174–180. <https://doi.org/10.1016/j.copsy.2019.03.005>
- Esterman, M., Noonan, S. K., Rosenberg, M., & DeGutis, J. (2013). In the zone or zoning out? Tracking behavioral and neural fluctuations during sustained attention. *Cerebral cortex*, 23(11), 2712–2723.
- Falkenberg, L. E., Specht, K., & Westerhausen, R. (2011). Attention and cognitive control networks assessed in a dichotic listening fMRI study. *Brain and Cognition*, 76(2), 276–285. <https://doi.org/10.1016/j.bandc.2011.02.006>
- Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*. *Journal of Cognitive Neuroscience*, 14(3), 340–347.
- Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, 4(December), 1–23. <https://doi.org/10.3389/fnhum.2010.00215>
- Feng, S., D'Mello, S., & Graesser, A. C. (2013). Mind wandering while reading easy and difficult texts. *Psychonomic Bulletin and Review*, 20(3), 586–592. <https://doi.org/10.3758/s13423-012-0367-y>
- Fiedler, L., Wöstmann, M., Graverson, C., Brandmeyer, A., Lunner, T., & Obleser, J. (2017). Single-channel in-ear-EEG detects the focus of auditory attention to concurrent tone streams and mixed speech. *Journal of Neural Engineering*, 14(3). <https://doi.org/10.1088/1741-2552/aa66dd>

- Formisano, E., Kim, D. S., Di Salle, F., Van de Moortele, P. F., Ugurbil, K., & Goebel, R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, 40(4), 859-869.
- Fortenbaugh, F. C., Degutis, J., & Esterman, M. (2017). Recent theoretical, neural, and clinical advances in sustained attention research. *Annals of the New York Academy of Sciences*, 1-22. <https://doi.org/10.1111/nyas.13318>
- Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2014). Shortened complex span tasks can reliably measure working memory capacity. *Memory and Cognition*, 43(2), 226-236. <https://doi.org/10.3758/s13421-014-0461-7>
- Foster, N. E. V., & Zatorre, R. J. (2010). Cortical structure predicts success in performing musical transformation judgments. *NeuroImage*, 53(1), 26-36. <https://doi.org/10.1016/j.neuroimage.2010.06.042>
- Foxe, J. J., & Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Frontiers in Psychology*, 2(JUL), 1-13. <https://doi.org/10.3389/fpsyg.2011.00154>
- François, C., Jaillet, F., Takerkar, S., & Schön, D. (2014). Faster sound stream segmentation in musicians than in non musicians. *PLoS ONE*, 9(7), 26-29. <https://doi.org/10.1371/journal.pone.0101340>
- Francois, C., & Schön, D. (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cerebral Cortex*, 21(10), 2357-2365. <https://doi.org/10.1093/cercor/bhr022>
- Frank, M. J. (2006). Hold your horses: A dynamic computational role for the subthalamic nucleus in decision making. *Neural Networks*, 19(8), 1120-1136. <https://doi.org/10.1016/j.neunet.2006.03.006>
- Franklin, M. S., Moore, K. S., Yip, C. Y., Jonides, J., Rattray, K., & Moher, J. (2008). The effects of musical training on verbal memory. *Psychology of Music*, 36(3), 353-365. <https://doi.org/10.1177/0305735607086044>
- Franklin, M. S., Smallwood, J., & Schooler, J. W. (2011). Catching the mind in flight: Using behavioral indices to detect mindless reading in real time. *Psychonomic Bulletin and Review*, 18(5), 992-997. <https://doi.org/10.3758/s13423-011-0109-6>
- Friedman, N. P., & Miyake, A. (2004). The Relations Among Inhibition and Interference Control Functions: A Latent-Variable Analysis. *Journal of Experimental Psychology: General*, 133(1), 101-135. <https://doi.org/10.1037/0096-3445.133.1.101>
- Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology Paris*, 100(1-3), 70-87. <https://doi.org/10.1016/j.jphysparis.2006.10.001>
- Fu, Q. J., & Galvin, J. J. (2007). Perceptual Learning and Auditory Training in Cochlear Implant Recipients. *Trends in Amplification*, 11(3), 193-205. <https://doi.org/10.1177/1084713807301379>
- Fuglsang, S. A., Dau, T., & Hjortkjær, J. (2017). Noise-robust cortical tracking of attended speech in real-world acoustic scenes. *NeuroImage*, 156(December 2016), 435-444. <https://doi.org/10.1016/j.neuroimage.2017.04.026>
- Fuller, C. D., Galvin, J. J., Maat, B., Free, R. H., & Başkent, D. (2014). The musician effect: Does it persist under degraded pitch conditions of cochlear implant simulations? *Frontiers in Neuroscience*, 8(8 JUN), 1-16. <https://doi.org/10.3389/fnins.2014.00179>
- Gaab, N., Gaser, C., Zaehle, T., Jancke, L., & Schlaug, G. (2003). Functional anatomy of pitch memory - An fMRI study with sparse temporal sampling. *NeuroImage*, 19(4), 1417-1426. [https://doi.org/10.1016/S1053-8119\(03\)00224-6](https://doi.org/10.1016/S1053-8119(03)00224-6)
- Gadea, M., & Espert, R. (2009). A comparison of the effects for sustained versus shifted attention on dichotic listening performance. *Laterality*, 14(3), 315-328. <https://doi.org/10.1080/13576500802440707>

- Gajęcki, T., & Nogueira, W. (2018). Deep learning models to remix music for cochlear implant users. *The Journal of the Acoustical Society of America*, 143(6), 3602–3615. <https://doi.org/10.1121/1.5042056>
- Gallun, F. J., Mason, C. R., & Kidd Jr, G. (2007). Task-dependent costs in processing. *Perception & Psychophysics*, 69(5), 757–771.
- Galuske, R. A. W., Schlote, W., Bratzke, H., & Singer, W. (2000). Interhemispheric asymmetries of the modular structure in human temporal cortex. *Science*, 289(5486), 1946–1949. <https://doi.org/10.1126/science.289.5486.1946>
- Gandolfi, E., Viterbori, P., Traverso, L., & Carmen Usai, M. (2014). Inhibitory processes in toddlers: A latent-variable approach. *Frontiers in Psychology*, 5(APR), 1–11. <https://doi.org/10.3389/fpsyg.2014.00381>
- Garavan, H., Ross, T. J., & Stein, E. A. (1999). Right hemispheric dominance of inhibitory control: An event-related functional MRI study. *Proceedings of the National Academy of Sciences of the United States of America*, 96(14), 8301–8306. <https://doi.org/10.1073/pnas.96.14.8301>
- Gartner, A., & Strobel, A. (2020). Individual differences in inhibitory control: A latent variable analysis. *Journal of Cognition*, 4(1), 1–18. <https://doi.org/10.5334/joc.150>
- Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *Journal of Neuroscience*, 23(27), 9240–9245. <https://doi.org/10.1523/jneurosci.23-27-09240.2003>
- Gaspelin, N., & Luck, S. J. (2018). The Role of Inhibition in Avoiding Distraction by Salient Stimuli. *Trends in Cognitive Sciences*, 22(1), 79–92. <https://doi.org/10.1016/j.tics.2017.11.001>
- Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 18(1), 1-16.
- Gaudrain, E., Grimault, N., Healy, E. W., & Béra, J. C. (2007). Effect of spectral smearing on the perceptual segregation of vowel sequences. *Hearing Research*, 231(1–2), 32–41. <https://doi.org/10.1016/j.heares.2007.05.001>
- Gaudrain, E., Grimault, N., Healy, E. W., & Béra, J.-C. (2008). Streaming of vowel sequences based on fundamental frequency in a cochlear-implant simulation. *The Journal of the Acoustical Society of America*, 124(5), 3076–3087. <https://doi.org/10.1121/1.2988289>
- Gavens, N., & Barrouillet, P. (2004). Delays of retention, processing efficiency, and attentional resources in working memory span development. *Journal of Memory and Language*, 51(4), 644–657.
- Gazzaley, A., & Nobre, A. (2012). Top-down modulation: Bridging selective attention and working memory. *Trends in Cognitive Sciences*, 16(2), 129–135. <https://doi.org/10.1016/j.tics.2011.11.014>. Top-down
- Geirnaert, S., Vandecappelle, S., Alickovic, E., De Cheveigne, A., Lalor, E., Meyer, B. T., Miran, S., Francart, T., & Bertrand, A. (2021). Electroencephalography-Based Auditory Attention Decoding: Toward Neurosteered Hearing Devices. *IEEE Signal Processing Magazine*, 38(4), 89–102. <https://doi.org/10.1109/MSP.2021.3075932>
- Geirnaert, S., Vandecappelle, S., Alickovic, E., de Cheveigné, A., Lalor, E., Meyer, B. T., Miran, S., Francart, T., & Bertrand, A. (2020). Neuro-steered hearing devices: Decoding auditory attention from the brain. *ArXiv*, 802895, 1–20.
- George, E. M., & Coch, D. (2011). Music training and working memory: An ERP study. *Neuropsychologia*, 49(5), 1083–1094. <https://doi.org/10.1016/j.neuropsychologia.2011.02.001>
- George, L., & Lécuyer, A. (2010). An overview of research on “passive” brain-computer interfaces for implicit human-computer interaction. *International Conference on Applied Bionics and Biomechanics ICABB 2010 - Workshop W1 “Brain-Computer Interfacing and Virtual Reality.”* <http://hal.inria.fr/docs/00/53/72/11/PDF/GeorgeL-LecuyerA.pdf>

- Gfeller, K., Witt, S., Adamek, M., Mehr, M., Rogers, J., Stordahl, J., & Ringgenberg, S. (2002). Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients. *Journal of the American Academy of Audiology*, 13(03), 132-145.
- Ghatan, P. H., Hsieh, J. C., Petersson, K. M., Stone-Elander, S., & Ingvar, M. (1998). Coexistence of attention-based facilitation and inhibition in the human cortex. *NeuroImage*, 7(1), 23-29. <https://doi.org/10.1006/nimg.1997.0307>
- Ghinescu, R., Schachtman, T. R., Ramsey, A. K., Gratton, G., & Fabiani, M. (2016). Conflict adaptation and cue competition during learning in an Eriksen flanker task. *PLoS ONE*, 11(12), 1-19. <https://doi.org/10.1371/journal.pone.0167119>
- Gillis, M., Canneyt, J. Van, Francart, T., & Vanthornhout, J. (2022). Neural tracking as a diagnostic tool to assess the auditory pathway 2 Marlies Gillis. *Hearing Research*, xxxx, 108607. <https://doi.org/10.1016/j.heares.2022.108607>
- Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, 15(4), 511-517. <https://doi.org/10.1038/nn.3063>
- Glendenning, K. K., & Masterton, R. B. (1998). Comparative morphometry of mammalian central auditory systems: variation in nuclei and form of the ascending system. *Brain, behavior and evolution*, 51(2), 59-89.
- Goehring, T., Keshavarzi, M., Carlyon, R. P., & Moore, B. C. J. (2019). Using recurrent neural networks to improve the perception of speech in non-stationary noise by people with cochlear implants. *The Journal of the Acoustical Society of America*, 146(1), 705-718. <https://doi.org/10.1121/1.5119226>
- Goldberg, J. M., & Brown, P. B. (1968). Functional organization of the dog superior olivary complex: an anatomical and electrophysiological study. *Journal of neurophysiology*, 31(4), 639-656.
- Golding, N. L., & Oertel, D. (2012). Synaptic integration in dendrites: Exceptional need for speed. *Journal of Physiology*, 590(22), 5563-5569. <https://doi.org/10.1113/jphysiol.2012.229328>
- Goll, J. C., Kim, L. G., Ridgway, G. R., Hailstone, J. C., Lehmann, M., Buckley, A. H., Crutch, S. J., & Warren, J. D. (2012). Impairments of auditory scene analysis in Alzheimer's disease. *Brain*, 135(1), 190-200. <https://doi.org/10.1093/brain/awr260>
- Gordon-Salant, S., & Cole, S. S. (2016). Effects of age and working memory capacity on speech recognition performance in noise among listeners with normal hearing. *Ear and Hearing*, 37(5), 593-602. <https://doi.org/10.1097/AUD.0000000000000316>
- Gordon-Salant, S., & Fitzgibbons, P. J. (2004). Effects of stimulus and noise rate variability on speech perception by younger and older adults. *The Journal of the Acoustical Society of America*, 115(4), 1808-1817. <https://doi.org/10.1121/1.1645249>
- Gordon, H. W. (1970). Hemispheric Asymmetries in the Perception of Musical Chords. *Cortex*, 6(4), 387-398. [https://doi.org/10.1016/S0010-9452\(70\)80003-X](https://doi.org/10.1016/S0010-9452(70)80003-X)
- Gordon, H. W. (1980). Degree of ear asymmetries for perception of dichotic chords and for illusory chord localization in musicians of different levels of competence. *Journal of Experimental Psychology: Human Perception and Performance*, 6(3), 516-527. <https://doi.org/10.1037/0096-1523.6.3.516>
- Gordon, H. W. (1978). Left Hemisphere Dominance for Rhythmic Elements in Dichotically-Presented Melodies. *Cortex*, 14(1), 58-70. [https://doi.org/10.1016/S0010-9452\(78\)80008-2](https://doi.org/10.1016/S0010-9452(78)80008-2)
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7(7 DEC), 1-13. <https://doi.org/10.3389/fnins.2013.00267>
- Grassi, M., Meneghetti, C., Toffalini, E., & Borella, E. (2017). Auditory and cognitive performance in elderly musicians and non musicians. *PLoS ONE*, 12(11), 1-21. <https://doi.org/10.1371/journal.pone.0187881>

- Greenberg, G. Z., & Larkin, W. D. (1968). Frequency-Response Characteristic of Auditory Observers Detecting Signals of a Single Frequency in Noise: The Probe-Signal Method. *The Journal of the Acoustical Society of America*, 44(6), 1513-1523.
- Greenlaw, K. M., Puschmann, S., & Coffey, E. B. J. (2020). Decoding of Envelope vs. Fundamental Frequency During Complex Auditory Stream Segregation. *Neurobiology of Language*, 1(3), 268-287. https://doi.org/10.1162/nol_a_00013
- Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2013). The musical stroop effect: Opening a new avenue to research on automatisms. *Experimental Psychology*, 60(4), 269-278. <https://doi.org/10.1027/1618-3169/a000197>
- Gregory, A. H. (1990). Listening to polyphonic music. *Psychology of Music*, 18(2), 163-170.
- Grimault, N., Bacon, S. P., & Micheyl, C. (2002). Auditory stream segregation on the basis of amplitude-modulation rate. *The Journal of the Acoustical Society of America*, 111(3), 1340-1348. <https://doi.org/10.1121/1.1452740>
- Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., & Collet, L. (2000). Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency. *The Journal of the Acoustical Society of America*, 108(1), 263-271. <https://doi.org/10.1121/1.429462>
- Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., & Collet, L. (2001). Perceptual auditory stream segregation of sequences of complex sounds in subjects with normal and impaired hearing. *British Journal of Audiology*, 35, 173-182.
- Gutschalk, A., Micheyl, C., & Oxenham, A. J. (2008). Neural correlates of auditory perceptual awareness under informational masking. *PLoS Biology*, 6(6), 1156-1165. <https://doi.org/10.1371/journal.pbio.0060138>
- Hackett, T. A. (2015). Anatomic organization of the auditory cortex. In *Handbook of Clinical Neurology* (1st ed., Vol. 129). Elsevier B.V. <https://doi.org/10.1016/B978-0-444-62630-1.00002-0>
- Hahn, B., Wolkenberg, F. A., Ross, T. J., Myers, C. S., Hershman, S. J., Stein, D. J., Kurup, P. K., & Stein, E. A. (2008). Divided versus selective attention: Evidence for common processing mechanisms. *Brain Research*, 1215, 137-146. <https://doi.org/10.1016/j.brainres.2008.03.058>
- Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision research*, 18(10), 1279-1296
- Hallion, L. S., Kusmierski, S. N., & Caulfield, M. K. (2020). Worry alters speed-accuracy tradeoffs but does not impair sustained attention. *Behaviour Research and Therapy*, 128(July 2019), 103597. <https://doi.org/10.1016/j.brat.2020.103597>
- Halpern, A., Dowling, W. J., & Bartlett, J. (1995). Bucknell Digital Commons Aging and experience in the recognition of musical transpositions Recognition of familiar and unfamiliar melodies in normal aging and Alzheimer's disease. *Psychology and Aging*, 325-342.
- Hanes, D. P., Patterson, W. F., & Schall, J. D. (1998). Role of frontal eye fields in countermanding saccades: visual, movement, and fixation activity. *Journal of neurophysiology*, 79(2), 817-834.
- Hannon, E. E., & Trainor, L. J. (2007). Music acquisition: effects of enculturation and formal training on development. *Trends in Cognitive Sciences*, 11(11), 466-472. <https://doi.org/10.1016/j.tics.2007.08.008>
- Hansen, J. C., & Hillyard, S. A. (1980). Endogeneous brain potentials associated with selective auditory attention. *Electroencephalography and clinical neurophysiology*, 49(3-4), 277-290.
- Hansen, M., Wallentin, M., & Vuust, P. (2012). Working memory and musical competence of musicians and non-musicians. *Psychology of Music*, 41(6), 779-793. <https://doi.org/10.1177/0305735612452186>

- Harding, E. E., Sammler, D., Henry, M. J., Large, E. W., & Kotz, S. A. (2019). Cortical tracking of rhythm in music and speech. *NeuroImage*, 185(August 2018), 96–101. <https://doi.org/10.1016/j.neuroimage.2018.10.037>
- Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S., & Engle, R. W. (2013). Working Memory Training May Increase Working Memory Capacity but Not Fluid Intelligence. *Psychological Science*, 24(12), 2409–2419. <https://doi.org/10.1177/0956797613492984>
- Hartmann, W. M., & Johnson, D. (1991). Stream segregation and peripheral channeling. *Music perception*, 9(2), 155-183.
- Hartmann, W. M., Mcadams, S., & Smith, B. K. (1990). Hearing a mistuned harmonic in an otherwise periodic complex tone. *Journal of the Acoustical Society of America*, 88(4), 1712–1724. <https://doi.org/10.1121/1.400246>
- Hasbroucq, T., & Guiard, Y. (1992). The effects of intensity and irrelevant location of a tactile stimulation in a choice reaction time task. *Neuropsychologia*, 30(1), 91-94.
- Hasher, L., & Zacks, R. T. (1988). Working Memory, Comprehension, and aging: a review and new view. *The Psychology of Learning*, 22, 193–225.
- Hasher, L., Lustig, C., & Zacks, R. (2007). Inhibitory mechanisms and the control of attention.
- Haslinger, B., Erhard, P., Altenmüller, E., Schroeder, U., Boecker, H., & Ceballos-Baumann, A. O. (2005). Transmodal sensorimotor networks during action observation in professional pianists. *Journal of Cognitive Neuroscience*, 17(2), 282–293. <https://doi.org/10.1162/0898929053124893>
- Haufe, S., Kim, J. W., Kim, I. H., Sonnleitner, A., Schrauf, M., Curio, G., & Blankertz, B. (2014). Electrophysiology-based detection of emergency braking intention in real-world driving. *Journal of Neural Engineering*, 11(5). <https://doi.org/10.1088/1741-2560/11/5/056011>
- Hausfeld, L., Disbergen, N. R., Valente, G., Zatorre, R. J., & Formisano, E. (2021). Modulating Cortical Instrument Representations During Auditory Stream Segregation and Integration With Polyphonic Music. *Frontiers in Neuroscience*, 15(September), 1–15. <https://doi.org/10.3389/fnins.2021.635937>
- Hausfeld, L., Riecke, L., Valente, G., & Formisano, E. (2018). Cortical tracking of multiple streams outside the focus of attention in naturalistic auditory scenes. *NeuroImage*, 181(March), 617–626. <https://doi.org/10.1016/j.neuroimage.2018.07.052>
- Hausfeld, L., Riecke, L., Valente, G., & Formisano, E. (2018). Cortical tracking of multiple streams outside the focus of attention in naturalistic auditory scenes. *NeuroImage*, 181(July), 617–626. <https://doi.org/10.1016/j.neuroimage.2018.07.052>
- Hawley, M. L., Litovsky, R. Y., & Culling, J. F. (2004). The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *The Journal of the Acoustical Society of America*, 115(2), 833–843. <https://doi.org/10.1121/1.1639908>
- He, J., Becic, E., Lee, Y.-C., & McCarley, J. S. (2011). Mind Wandering Behind the Wheel. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 53(1), 13–21. <https://doi.org/10.1177/0018720810391530>
- Hee Lee, J., & Humes, L. E. (2012). Effect of fundamental-frequency and sentence-onset differences on speech-identification performance of young and older adults in a competing-talker background. *The Journal of the Acoustical Society of America*, 132(3), 1700–1717. <https://doi.org/10.1121/1.4740482>
- Helfert, R. H., Snead, C. R., & Altschuler, R. A. (1991). The ascending auditory pathways. *Neurobiology of hearing: the central auditory system*, 1-25.
- Helge Johnsen, B., Christian Laberg, J., Eid, J., & Hugdahl, K. (2002). Dichotic listening and sleep deprivation: Vigilance effects. *Scandinavian Journal of Psychology*, 43(5), 413–417. <https://doi.org/10.1111/1467-9450.00309>

- Helton, W. S. (2009). Impulsive responding and the sustained attention to response task. *Journal of Clinical and Experimental Neuropsychology*, 31(1), 39–47. <https://doi.org/10.1080/13803390801978856>
- Helton, W. S., & Russell, P. N. (2011). Feature absence-presence and two theories of lapses of sustained attention. *Psychological Research*, 75(5), 384–392. <https://doi.org/10.1007/s00426-010-0316-1>
- Herff, S. A., & Czernochowski, D. (2019). The role of divided attention and expertise in melody recognition. *Musicae Scientiae*, 23(1), 69–86. <https://doi.org/10.1177/1029864917731126>
- Herholz, S. C., & Zatorre, R. J. (2012). Musical Training as a Framework for Brain Plasticity: Behavior, Function, and Structure. *Neuron*, 76(3), 486–502. <https://doi.org/10.1016/j.neuron.2012.10.011>
- Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, 4(4), 131–138.
- Hilbert, S., Nakagawa, T., Puci, P., Zech, A., & Böhner, M. (2015). The digit span backwards task: Verbal and visual cognitive strategies in working memory assessment. *European Journal of Psychological Assessment*, 31(3), 174–180.
- Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology*, 3(APR), 1–14. <https://doi.org/10.3389/fpsyg.2012.00096>
- Holden, L. K., Finley, C. C., Firszt, J. B., Timothy, A., Brenner, C., Potts, L. G., Gotter, B. D., Vanderhoof, S. S., Mispagel, K., & Heydebrand, G. (2013). Factors Affecting Open-Set Word Recognition in Adults with Cochlear Implants. *Ear and Hearing*, 34(3), 342–360. <https://doi.org/10.1097/AUD.0b013e3182741aa7>. Factors
- Holdgraf, C. R., Rieger, J. W., Micheli, C., Martin, S., Knight, R. T., & Theunissen, F. E. (2017). Encoding and Decoding Models in Cognitive Electrophysiology. *Frontiers in Systems Neuroscience*, 11(September). <https://doi.org/10.3389/fnsys.2017.00061>
- Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behavior and Brain Sciences*, 9, 1–66.
- Hommel, B. (1993). The role of attention for the Simon effect. *Psychological Research*, 55(3), 208–222. <https://doi.org/10.1007/BF00419608>
- Hommel, B. (2011). The Simon effect as tool and heuristic. *Acta Psychologica*, 136(2), 189–202. <https://doi.org/10.1016/j.actpsy.2010.04.011>
- Hommel, B., Chapman, C. S., Cisek, P., Neyedli, H. F., Song, J. H., & Welsh, T. N. (2019). No one knows what attention is. *Attention, Perception, and Psychophysics*, 81(7), 2288–2303. <https://doi.org/10.3758/s13414-019-01846-w>
- Hommet, C., Mondon, K., Berrut, G., Gouyer, Y., Isingrini, M., Constans, T., & Belzung, C. (2010). Central auditory processing in aging: The dichotic listening paradigm. *Journal of Nutrition, Health and Aging*, 14(9), 751–756. <https://doi.org/10.1007/s12603-010-0097-7>
- Hou, J., Chen, C., Wang, Y., Liu, Y., He, Q., Jin, L., & Dong, Q. (2014). Superior pitch identification ability is associated with better mental rotation performance. *Psychomusicology: Music, Mind, and Brain*, 24(2), 136–146. <https://doi.org/10.1177/10298649211013409>
- Howard, S. J., Johnson, J., & Pascual-Leone, J. (2014). Clarifying inhibitory control: Diversity and development of attentional inhibition. *Cognitive Development*, 31(1), 1–21. <https://doi.org/10.1016/j.cogdev.2014.03.001>
- Hugdahl, K., Westerhausen, R., Alho, K., Medvedev, S., Laine, M., & Hämäläinen, H. (2009). Attention and cognitive control: Unfolding the dichotic listening story: Cognition and Neurosciences. *Scandinavian Journal of Psychology*, 50(1), 11–22. <https://doi.org/10.1111/j.1467-9450.2008.00676.x>
- Hughes, R. W., Vachon, F., & Jones, D. M. (2007). Disruption of Short-Term Memory by Changing and Deviant Sounds: Support for a Duplex-Mechanism Account of Auditory Distraction. *Journal of Experimental*

- Psychology: Learning Memory and Cognition, 33(6), 1050–1061. <https://doi.org/10.1037/0278-7393.33.6.1050>
- Humes, L. E., Halling, D., & Coughlin, M. (1996). Reliability and stability of various hearing-aid outcome measures in a group of elderly hearing-aid wearers. *Journal of Speech, Language, and Hearing Research*, 39(5), 923–935. <https://doi.org/10.1044/jshr.3905.923>
- Humphries, C., Liebenthal, E., & Binder, J. R. (2010). Tonotopic organization of human auditory cortex. *NeuroImage*, 50(3), 1202–1211. <https://doi.org/10.1016/j.neuroimage.2010.01.046>
- Hutchinson, S., Lee, L. H. L., Gaab, N., & Schlaug, G. (2003). Cerebellar volume of musicians. *Cerebral Cortex*, 13(9), 943–949. <https://doi.org/10.1093/cercor/13.9.943>
- Hutsler, J. J., & Gazzaniga, M. S. (1996). Acetylcholinesterase staining in human auditory and language cortices: Regional variation of structural features. *Cerebral Cortex*, 6(2), 260–270. <https://doi.org/10.1093/cercor/6.2.260>
- Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., & Schlaug, G. (2009). The effects of musical training on structural brain development: A longitudinal study. *Annals of the New York Academy of Sciences*, 1169, 182–186. <https://doi.org/10.1111/j.1749-6632.2009.04852.x>
- Hyrskykari, A. (2006). Eyes in Attentive Interfaces: Experiences from Creating iDict, A Gaze-Aware Reading Aid (Issue January 2006). <http://acta.uta.fi>
- Ihlefeld, A., & Shinn-Cunningham, B. (2008). Spatial release from energetic and informational masking in a selective speech identification task. *The Journal of the Acoustical Society of America*, 123(6), 4369–4379. <https://doi.org/10.1121/1.2904826>
- Ikkai, A., & Curtis, C. E. (2011). Common neural mechanisms supporting spatial working memory, attention and motor intention. *Neuropsychologia*, 49(6), 1428–1434. <https://doi.org/10.1016/j.neuropsychologia.2010.12.020>
- 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>
- Iverson, P. (1995). Auditory stream segregation by musical timbre: effects of static and dynamic acoustic attributes. *Journal of Experimental Psychology: Human Perception and Performance*, 21(4), 751.
- Jackson, J. D., & Balota, D. A. (2012). Mind-wandering in Younger and Older Adults: Converging Evidence from the Sustained Attention to Response Task and Reading for Comprehension. *Psychology and Aging*, 27(1), 106–119. <https://doi.org/10.1037/a0023933>. Mind-wandering
- Jacob, R. J. K. (1990). WHAT YOU LOOK AT IS WHAT YOU GET: EYE MOVEMENT-BASED INTERACTION TECHNIQUES. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, March 1990, 11–18.
- Jaeggi, S. M., Buschkuhl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, 18(4), 394–412. <https://doi.org/10.1080/09658211003702171>
- Jain, R., & Shore, S. (2006). External inferior colliculus integrates trigeminal and acoustic information: unit responses to trigeminal nucleus and acoustic stimulation in the guinea pig. *Neuroscience letters*, 395(1), 71–75.
- Jakobson, L. S., Lewycky, S. T., Kilgour, A. R., & Stoesz, B. M. (2008). Memory for verbal and visual material in highly trained musicians. *Music Perception*, 26(1), 41–55. <https://doi.org/10.1525/mp.2008.26.1.41>
- Jalbert, A., Neath, I., Bireta, T. J., & Surprenant, A. M. (2011). When does length cause the word length effect?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(2), 338.

- James, P. J., Krishnan, S., & Aydelott, J. (2014). Working memory predicts semantic comprehension in dichotic listening in older adults. *Cognition*, 133(1), 32–42. <https://doi.org/10.1016/j.cognition.2014.05.014>
- Janata, P., Tillmann, B., & Bharucha, J. J. (2002). Listening to polyphonic music recruits domain-general attention and working memory circuits. *Cognitive, Affective and Behavioral Neuroscience*, 2(2), 121–140. <https://doi.org/10.3758/CABN.2.2.121>
- Jäncke, L. (2009). Music drives brain plasticity. *F1000 Biology Reports*, 1(October), 1–6. <https://doi.org/10.3410/b1-78>
- Jäncke, L., Specht, K., Shah, J. N., & Hugdahl, K. (2003). Focused attention in a simple dichotic listening task: An fMRI experiment. *Cognitive Brain Research*, 16(2), 257–266. [https://doi.org/10.1016/S0926-6410\(02\)00281-1](https://doi.org/10.1016/S0926-6410(02)00281-1)
- Janse, E. (2012). A non-auditory measure of interference predicts distraction by competing speech in older adults. *Aging, Neuropsychology, and Cognition*, 19(6), 741–758. <https://doi.org/10.1080/13825585.2011.652590>
- Jerde, T. A., Childs, S. K., Handy, S. T., Nagode, J. C., & Pardo, J. V. (2011). Dissociable systems of working memory for rhythm and melody. *NeuroImage*, 57(4), 1572–1579. <https://doi.org/10.1016/j.neuroimage.2011.05.061>
- Jerger, S., Martin, R. C., & Pirozzolo, F. J. (1988). A developmental study of the auditory stroop effect. *Brain and Language*, 35(1), 86–104. [https://doi.org/10.1016/0093-934X\(88\)90102-2](https://doi.org/10.1016/0093-934X(88)90102-2)
- 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. *Psychological Science*, 24(10), 1995–2004. <https://doi.org/10.1177/0956797613482467>
- Jones, M. R. (1976). Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychological review*, 83(5), 323.
- Jones, M. R., & Yee, W. (1997). Sensitivity to time change: The role of context and skill. *Journal of Experimental Psychology: Human Perception and Performance*, 23(3), 693.
- Jones, E. G. (2003). Chemically defined parallel pathways in the monkey auditory system. *Annals of the New York Academy of Sciences*, 999(1), 218–233.
- Jonides, J., Schumacher, E. H., Smith, E. E., Lauber, E. J., Awh, E., Minoshima, S., & Koeppe, R. A. (1997). Verbal working memory load affects regional brain activation as measured by PET. *Journal of Cognitive Neuroscience*, 9(4), 462–475. <https://doi.org/10.1162/jocn.1997.9.4.462>
- Kaganovich, N., Kim, J., Herring, C., Schumaker, J., Macpherson, M., & Weber-Fox, C. (2013). Musicians show general enhancement of complex sound encoding and better inhibition of irrelevant auditory change in music: An ERP study. *European Journal of Neuroscience*, 37(8), 1295–1307. <https://doi.org/10.1111/ejn.12110>
- Kaiser, J. (2015). Dynamics of auditory working memory. *Frontiers in Psychology*, 6(MAY), 1–6. <https://doi.org/10.3389/fpsyg.2015.00613>
- Kam, J. W. Y., & Handy, T. C. (2013). The neurocognitive consequences of the wandering mind: a mechanistic account of sensory-motor decoupling. *Frontiers in Psychology*, 4(October), 1–13. <https://doi.org/10.3389/fpsyg.2013.00725>
- Kam, J. W. Y., Dao, E., Farley, J., Fitzpatrick, K., Smallwood, J., Schooler, J. W., & Handy, T. C. (2011). Slow fluctuations in attentional control of sensory cortex. *Journal of Cognitive Neuroscience*, 23(2), 460–470. <https://doi.org/10.1162/jocn.2010.21443>

- Kam, J. W. Y., Xu, J., & Handy, T. C. (2014). I don't feel your pain (as much): The desensitizing effect of mind wandering on the perception of others' discomfort. *Cognitive, Affective and Behavioral Neuroscience*, 14(1), 286–296. <https://doi.org/10.3758/s13415-013-0197-z>
- Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). Variation in Working Memory Capacity as Variation in Executive Attention and Control. In *Variation in Working Memory*. <https://doi.org/10.1093/acprof:oso/9780195168648.003.0002>
- Kane, M. J., & Engle, R. W. (2003). Working-Memory Capacity and the Control of Attention: The Contributions of Goal Neglect, Response Competition, and Task Set to Stroop Interference. *Journal of Experimental Psychology: General*, 132(1), 47–70. <https://doi.org/10.1037/0096-3445.132.1.47>
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of experimental psychology: General*, 133(2), 189.
- Kane, M. J., Meier, M. E., Smeekens, B. A., Gross, G. M., Chun, C. A., Silvia, P. J., & Kwapil, T. R. (2016). Individual Differences In The Executive Control Of Attention, Memory, And Thought, And Their Associations With Schizotypy. *Physiology & Behavior*, 145(8), 1017–1048. <https://doi.org/10.1037/xge0000184>. Individual
- Kane, M. J., Smeekens, B. A., Meier, M. E., Welhaf, M. S., & Phillips, N. E. (2021). Testing the construct validity of competing measurement approaches to probed mind-wandering reports. In *Behavior Research Methods* (Vol. 53, Issue 6). *Behavior Research Methods*. <https://doi.org/10.3758/s13428-021-01685-4>
- Kappenman, E. S., & Luck, S. J. (2010). The effects of electrode impedence on data quality and statistical significance in ERP recordings. *Psychophysiology*, 47(5), 888–904. <https://doi.org/10.1111/j.1469-8986.2010.01009.x>
- Kastner, S., & Ungerleider, L. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, 23, 315–341. <http://www.annualreviews.org/doi/abs/10.1146/annurev.neuro.23.1.315>
- Kenna, M. A. (2015). Acquired hearing loss in children. *Otolaryngologic Clinics of North America*, 48(6), 933–953.
- Kerber, S., & Seeber, B. U. (2012). Sound localization in noise by normal-hearing listeners and cochlear implant users. *Ear and Hearing*, 33(4), 445–457. <https://doi.org/10.1097/AUD.0b013e318257607b>
- Kern, P., Heilbron, M., Lange, F. P. De, & Spaak, E. (2022). Cortical activity during naturalistic music listening reflects short-range predictions based on long-term experience. *BioRxiv Preprint*.
- Kidd, G., Best, V., & Mason, C. R. (2008). Listening to every other word: Examining the strength of linkage variables in forming streams of speech. *The Journal of the Acoustical Society of America*, 124(6), 3793–3802. <https://doi.org/10.1121/1.2998980>
- Kiefer, M., Ahlegian, M., & Spitzer, M. (2005). Working memory capacity, indirect semantic priming, and stroop interference: Pattern of interindividual prefrontal performance differences in healthy volunteers. *Neuropsychology*, 19(3), 332–344. <https://doi.org/10.1037/0894-4105.19.3.332>
- King, A. J., & Walker, K. M. (2020). Listening in complex acoustic scenes. *Current Opinion in Physiology*, 18, 63–72. <https://doi.org/10.1016/j.cophys.2020.09.001>
- Kishon-Rabin, L., Amir, O., Vexler, Y., & Zaltz, Y. (2001). Pitch discrimination: Are professional musicians better than non-musicians? *Journal of Basic and Clinical Physiology and Pharmacology*, 12(2), 125–144. <https://doi.org/10.1515/JBCPP.2001.12.2.125>
- 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(MAR), 1–21. <https://doi.org/10.3389/fpsyg.2017.00230>

- 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>
- Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., & Gruber, O. (2009). Functional architecture of verbal and tonal working memory: An fMRI study. *Human Brain Mapping*, 30(3), 859–873. <https://doi.org/10.1002/hbm.20550>
- Kohlberg, G. D., Mancuso, D. M., Chari, D. A., & Lalwani, A. K. (2015). Music Engineering as a Novel Strategy for Enhancing Music Enjoyment in the Cochlear Implant Recipient. *Behavioural Neurology*, 2015. <https://doi.org/10.1155/2015/829680>
- KOHLER, D. W. (1947). Gestalt psychology.
- Kondo, H. M., Van Loon, A. M., Kawahara, J. I., & Moore, B. C. J. (2017). Auditory and visual scene analysis: An overview. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1714). <https://doi.org/10.1098/rstb.2016.0099>
- Kostyrka-Allchorne, K., Cooper, N. R., & Simpson, A. (2017). The relationship between television exposure and children's cognition and behaviour: A systematic review. *Developmental Review*, 44, 19–58. <https://doi.org/10.1016/j.dr.2016.12.002>
- Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, 11(8), 599–605. <https://doi.org/10.1038/nrn2882>
- Kubanek, J., Brunner, P., Gunduz, A., Poeppel, D., & Schalk, G. (2013). The Tracking of Speech Envelope in the Human Cortex. *PLoS ONE*, 8(1). <https://doi.org/10.1371/journal.pone.0053398>
- Kubanek, J., Brunner, P., Gunduz, A., Poeppel, D., & Schalk, G. (2013). The Tracking of Speech Envelope in the Human Cortex. *PLoS ONE*, 8(1). <https://doi.org/10.1371/journal.pone.0053398>
- Kumagai, Y., Arvaneh, M., & Tanaka, T. (2017). Familiarity affects entrainment of EEG in music listening. *Frontiers in Human Neuroscience*, 11(July), 1–8. <https://doi.org/10.3389/fnhum.2017.00384>
- Kumagai, Y., Matsui, R., & Tanaka, T. (2018). Music familiarity affects EEG entrainment when little attention is paid. *Frontiers in Human Neuroscience*, 12(November), 1–11. <https://doi.org/10.3389/fnhum.2018.00444>
- Kuriki, S., Kanda, S., & Hirata, Y. (2006). Effects of musical experience on different components of meg responses elicited by sequential piano-tones and chords. *Journal of Neuroscience*, 26(15), 4046–4053. <https://doi.org/10.1523/JNEUROSCI.3907-05.2006>
- Labar, K. S., Gitelman, D. R., Parrish, T. B., & Mesulam, M. M. (1999). Neuroanatomic overlap of working memory and spatial attention networks: A functional MRI comparison within subjects. *NeuroImage*, 10(6), 695–704. <https://doi.org/10.1006/nimg.1999.0503>
- Lachaux, J. P., Rudrauf, D., & Kahane, P. (2003). Intracranial EEG and human brain mapping. *Journal of Physiology Paris*, 97(4–6), 613–628. <https://doi.org/10.1016/j.jphysparis.2004.01.018>
- Lackner, J. R., & Goldstein, L. M. (1974). Primary auditory stream segregation of repeated consonant—vowel sequences. *The Journal of the Acoustical Society of America*, 56(5), 1651–1652.
- Lalor, E. C., Power, A. J., Reilly, R. B., & Foxe, J. J. (2009). Resolving precise temporal processing properties of the auditory system using continuous stimuli. *Journal of Neurophysiology*, 102(1), 349–359. <https://doi.org/10.1152/jn.90896.2008>
- Lambez, B., Harwood-Gross, A., Golumbic, E. Z., & Rasseovsky, Y. (2020). Non-pharmacological interventions for cognitive difficulties in ADHD: A systematic review and meta-analysis. *Journal of Psychiatric Research*, 120(October 2019), 40–55. <https://doi.org/10.1016/j.jpsychires.2019.10.007>

- Lange, K., Kühn, S., & Filevich, E. (2015). "Just another tool for online studies" (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLoS ONE*, 10(6), 1–14. <https://doi.org/10.1371/journal.pone.0130834>
- Langers, D. R. M., & Van Dijk, P. (2012). Mapping the tonotopic organization in human auditory cortex with minimally salient acoustic stimulation. *Cerebral Cortex*, 22(9), 2024–2038. <https://doi.org/10.1093/cercor/bhr282>
- Lappin, J. S., & Eriksen, C. W. (1966). Use of a delayed signal to stop a visual reaction-time response. *Journal of Experimental Psychology*, 72(6), 805–811. <https://doi.org/10.1037/h0021266>
- Larson, E., & Lee, A. K. C. (2013). Influence of preparation time and pitch separation in switching of auditory attention between streams. *The Journal of the Acoustical Society of America*, 134(2), EL165–EL171. <https://doi.org/10.1121/1.4812439>
- Laszlo, S., Ruiz-Blondet, M., Khalifian, N., Chu, F., & Jin, Z. (2014). A direct comparison of active and passive amplification electrodes in the same amplifier system. *Journal of Neuroscience Methods*, 235, 298–307. <https://doi.org/10.1016/j.jneumeth.2014.05.012>
- Lavie, N., & De Fockert, J. W. (2005). The role of working memory in attentional capture. *Psychonomic Bulletin and Review*, 12(4), 669–674.
- Lawo, V., Fels, J., Oberem, J., & Koch, I. (2014). Intentional attention switching in dichotic listening: Exploring the efficiency of nonspatial and spatial selection. *Quarterly Journal of Experimental Psychology*, 67(10), 2010–2024. <https://doi.org/10.1080/17470218.2014.898079>
- Le, M., & Kerangal, G. De. (2020). The effect of ageing on auditory scene analysis : evidence from a change detection paradigm. June.
- Lécuyer, A., Lotte, F., Reilly, R. B., & Colledge, T. (2008). *Brain-Computer Interfaces, Virtual Reality, and Videogames*.
- Lee, A. K. C., & Shinn-Cunningham, B. G. (2008). Effects of frequency disparities on trading of an ambiguous tone between two competing auditory objects. *The Journal of the Acoustical Society of America*, 123(6), 4340–4351. <https://doi.org/10.1121/1.2908282>
- Lee, Y. shiow, Lu, M. ju, & Ko, H. ping. (2007). Effects of skill training on working memory capacity. *Learning and Instruction*, 17(3), 336–344. <https://doi.org/10.1016/j.learninstruc.2007.02.010>
- Lehto, J., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology*, 21(1), 59–80.
- Leigh, J. R., Moran, M., Hollow, R., & Dowell, R. C. (2016). Evidence-based guidelines for recommending cochlear implantation for postlingually deafened adults. *International Journal of Audiology*, 55(March), S3–S8. <https://doi.org/10.3109/14992027.2016.1146415>
- Lerousseau, J. P., Hidalgo, C., & Schön, D. (2020). Musical training for auditory rehabilitation in hearing loss. *Journal of Clinical Medicine*, 9(4). <https://doi.org/10.3390/jcm9041058>
- Liégeois-Chauvel, C., De Graaf, J. B., Laguitton, V., & Chauvel, P. (1999). Specialization of left auditory cortex for speech perception in man depends on temporal coding. *Cerebral Cortex*, 9(5), 484–496. <https://doi.org/10.1093/cercor/9.5.484>
- Lieu, J. E. C., Kenna, M., Anne, S., & Davidson, L. (2020). Hearing Loss in Children: A Review. *JAMA - Journal of the American Medical Association*, 324(21), 2195–2205. <https://doi.org/10.1001/jama.2020.17647>
- Limb, C. J., & Roy, A. T. (2014). Technological, biological, and acoustical constraints to music perception in cochlear implant users. *Hearing research*, 308, 13–26.
- Lipp, R., Kitterick, P., Summerfield, Q., Bailey, P. J., & Paul-Jordanov, I. (2010). Concurrent sound segregation based on inharmonicity and onset asynchrony. *Neuropsychologia*, 48(5), 1417–1425. <https://doi.org/10.1016/j.neuropsychologia.2010.01.009>

- Lo, C. C., Boucher, L., Paré, M., Schall, J. D., & Wang, X. J. (2009). Proactive inhibitory control and attractor dynamics in countermanding action: A spiking neural circuit model. *Journal of Neuroscience*, 29(28), 9059–9071. <https://doi.org/10.1523/JNEUROSCI.6164-08.2009>
- Logan, G. D. (1981). Attention, automaticity, and the ability to stop a speeded choice response. *Attention and performance IX*, 205-222.
- Logan, G. D., Van Zandt, T., Verbruggen, F., & Wagenmakers, E. J. (2014). On the ability to inhibit thought and action: General and special theories of an act of control. *Psychological Review*, 121(1), 66–95. <https://doi.org/10.1037/a0035230>
- Logie, R. H., Zucco, G. M., & Baddeley, A. D. (1990). Interference with visual short-term memory. *Acta psychologica*, 75(1), 55-74.
- Logie, R. H. (2016). Retiring the central executive. *Quarterly Journal of Experimental Psychology*, 69(10), 2093–2109. <https://doi.org/10.1080/17470218.2015.1136657>
- Loizou, P. C., Hu, Y., Litovsky, R., Yu, G., Peters, R., Lake, J., & Roland, P. (2009). Speech recognition by bilateral cochlear implant users in a cocktail-party setting. *The Journal of the Acoustical Society of America*, 125(1), 372–383. <https://doi.org/10.1121/1.3036175>
- Long, D. L., & Prat, C. S. (2002). Working memory and stroop interference: An individual differences investigation. *Memory and Cognition*, 30(2), 294–301. <https://doi.org/10.3758/BF03195290>
- Looi, V., Gfeller, K., & Driscoll, V. (2012). Music appreciation and training for cochlear implant recipients: A review. *Seminars in Hearing*, 33(4), 307–334. <https://doi.org/10.1055/s-0032-1329222>
- Lopes da Silva, F. H., & Van Rotterdam, A. (2005). Biophysical aspects of EEG and magnetoencephalographic generation. *Electroencephalography: basic principles, clinical applications and related fields*, 5th edn. Lippincott, Williams & Wilkins, New York.
- Lopes da Silva FH (2002) Electrical potentials. In: Ramachandran VS (ed) *Encyclopedia of the human brain*. Elsevier, New York, pp 147–67
- Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B. C. J. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences of the United States of America*, 103(49), 18866–18869. <https://doi.org/10.1073/pnas.0607364103>
- Loui, P., & Wessel, D. (2007). Harmonic expectation and affect in Western music: Effects of attention and training. *Perception and Psychophysics*, 69(7), 1084–1092. <https://doi.org/10.3758/BF03193946>
- Lu, K., Xu, Y., Yin, P., Oxenham, A. J., Fritz, J. B., & Shamma, S. A. (2017). Temporal coherence structure rapidly shapes neuronal interactions. *Nature Communications*, 8. <https://doi.org/10.1038/ncomms13900>
- Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.
- Luiz, C. B., Camargo, N. V. De, Habiro, J., Miguel, D. S., & Gil, D. (2021). Auditory Abilities in Individuals With and Without Formal Musical Training. *Journal of Hearing Science*, 11(3), 27–31. <https://doi.org/10.17430/jhs.2021.11.3.3>
- Lumani, A., & Zhang, H. (2010). Responses of neurons in the rat's dorsal cortex of the inferior colliculus to monaural tone bursts. *Brain research*, 1351, 115-129.
- Luo, C., Guo, Z. wei, Lai, Y. xiu, Liao, W., Liu, Q., Kendrick, K. M., Yao, D. zhong, & Li, H. (2012). Musical training induces functional plasticity in perceptual and motor networks: Insights from resting-state fMRI. *PLoS ONE*, 7(5), 1–10. <https://doi.org/10.1371/journal.pone.0036568>
- Macdonald, J. S. P., Mathan, S., & Yeung, N. (2011). Trial-by-trial variations in subjective attentional state are reflected in ongoing prestimulus EEG alpha oscillations. *Frontiers in Psychology*, 2(MAY), 1–16. <https://doi.org/10.3389/fpsyg.2011.00082>

- Macherey, O., & Carlyon, R. P. (2014). Cochlear implants. *Current Biology*, 24(18), R878–R884. <https://doi.org/10.1016/j.cub.2014.06.053>
- Macken, W. J., Tremblay, S., Houghton, R. J., Nicholls, A. P., & Jones, D. M. (2003). Does Auditory Streaming Require Attention? Evidence From Attentional Selectivity in Short-Term Memory. *Journal of Experimental Psychology: Human Perception and Performance*, 29(1), 43–51. <https://doi.org/10.1037/0096-1523.29.1.43>
- Mackie, M. A., Van Dam, N. T., & Fan, J. (2013). Cognitive control and attentional functions. *Brain and Cognition*, 82(3), 301–312. <https://doi.org/10.1016/j.bandc.2013.05.004>
- MacLeod, C. M. (1991). Half a century of research on the stroop effect: An integrative review. *Psychological Bulletin*, 109(2), 163–203. <https://doi.org/10.1037/0033-2909.109.2.163>
- Maddox, R. K., & Shinn-Cunningham, B. G. (2012). Influence of task-relevant and task-irrelevant feature continuity on selective auditory attention. *JARO - Journal of the Association for Research in Otolaryngology*, 13(1), 119–129. <https://doi.org/10.1007/s10162-011-0299-7>
- Madsen, J., Margulis, E. H., Simchy-Gross, R., & Parra, L. C. (2019). Music synchronizes brainwaves across listeners with strong effects of repetition, familiarity and training. *Scientific Reports*, 9(1), 1–8. <https://doi.org/10.1038/s41598-019-40254-w>
- Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: Behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience*, 18(2), 199–211. <https://doi.org/10.1162/jocn.2006.18.2.199>
- Mallet, N., Schmidt, R., Leventhal, D., Chen, F., Amer, N., Boraud, T., & Berke, J. D. (2016). Arkypallidal Cells Send a Stop Signal to Striatum. *Neuron*, 89(2), 308–316. <https://doi.org/10.1016/j.neuron.2015.12.017>
- Manning, F. C., Harris, J., & Schutz, M. (2017). Temporal prediction abilities are mediated by motor effector and rhythmic expertise. *Experimental Brain Research*, 235(3), 861–871. <https://doi.org/10.1007/s00221-016-4845-8>
- Manning, F. C., & Schutz, M. (2015). Trained to keep a beat: movement-related enhancements to timing perception in percussionists and non-percussionists. *Psychological Research*, 80(4), 532–542. <https://doi.org/10.1007/s00426-015-0678-5>
- Manning, F. C., Siminoski, A., & Schutz, M. (2020). Exploring the effects of effectors: Finger synchronization aids rhythm perception similarly in both pianists and non-pianists. *Music Perception*, 37(3), 196–207. <https://doi.org/10.1525/MP.2020.37.3.196>
- Marazita, M. L., Ploughman, L. M., Rawlings, B., Remington, E., Arnos, K. S., & Nance, W. E. (1993). Genetic epidemiological studies of early-onset deafness in the U.S. school-age population. *American Journal of Medical Genetics*, 46(5), 486–491. <https://doi.org/10.1002/ajmg.1320460504>
- Marion, G., Di Liberto, G. M., & Shamma, S. A. (2021). The music of silence. Part I: Responses to musical imagery encode melodic expectations and acoustics. *Journal of Neuroscience*, 41(35), 7435–7448. <https://doi.org/10.1523/JNEUROSCI.0183-21.2021>
- Marozeau, J., Innes-Brown, H., & Blamey, P. (2013). THE EFFECT OF TIMBRE AND LOUDNESS ON MELODY SEGREGATION. *Music Percep*, 30(3), 259–274.
- Marozeau, J., Innes-Brown, H., Grayden, D. B., Burkitt, A. N., & Blamey, P. J. (2010). The effect of visual cues on auditory stream segregation in musicians and non-musicians. *PLoS ONE*, 5(6), 1–10. <https://doi.org/10.1371/journal.pone.0011297>
- Mathewson, K. E., Harrison, T. J. L., & Kizuk, S. A. D. (2017). High and dry? Comparing active dry EEG electrodes to active and passive wet electrodes. *Psychophysiology*, 54(1), 74–82. <https://doi.org/10.1111/psyp.12536>

- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44(2), 314–324. <https://doi.org/10.3758/s13428-011-0168-7>
- Matthews, T. E., Thibodeau, J. N. L., Gunther, B. P., & Penhune, V. B. (2016). The Impact of Instrument-Specific Musical Training on Rhythm Perception and Production. *Frontiers in Psychology*, 7(February), 1–16. <https://doi.org/10.3389/fpsyg.2016.00069>
- McAdams, S. (2013). Timbre as a structuring force in music. *Proceedings of Meetings on Acoustics*, 19. <https://doi.org/10.1121/1.4799391>
- McAdams, S., & Matzkin, D. (2003). The Roots of Musical Variation in Perceptual Similarity and Invariance. *The Cognitive Neuroscience of Music*, 1–15. <https://doi.org/10.1093/acprof:oso/9780198525202.003.0006>
- McCloy, D. R., Lau, B. K., Larson, E., Pratt, K. A. I., & Lee, A. K. C. (2017). Pupillometry shows the effort of auditory attention switching. *The Journal of the Acoustical Society of America*, 141(4), 2440–2451. <https://doi.org/10.1121/1.4979340>
- McDermott, H. J. (2004). Music Perception with Cochlear Implants: A Review. *Trends in Amplification*, 8(2), 49–82. <https://doi.org/10.1177/108471380400800203>
- McDermott, J. H., Wroblewski, D., & Oxenham, A. J. (2011). Recovering sound sources from embedded repetition. *Proceedings of the National Academy of Sciences of the United States of America*, 108(3), 1188–1193. <https://doi.org/10.1073/pnas.1004765108>
- McNab, F., Leroux, G., Strand, F., Thorell, L., Bergman, S., & Klingberg, T. (2008). Common and unique components of inhibition and working memory: An fMRI, within-subjects investigation. *Neuropsychologia*, 46(11), 2668–2682. <https://doi.org/10.1016/j.neuropsychologia.2008.04.023>
- McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension. *Journal of Experimental Psychology General*, 141(2), 302–320. <https://doi.org/10.1037/a0025250>
- Medina, D., & Barraza, P. (2019). Efficiency of attentional networks in musicians and non-musicians. *Heliyon*, 5(3). <https://doi.org/10.1016/j.heliyon.2019.e01315>
- Meier, M. E., & Kane, M. J. (2013). Working memory capacity and stroop interference: Global versus local indices of executive control. *Journal of Experimental Psychology: Learning Memory and Cognition*, 39(3), 748–759. <https://doi.org/10.1037/a0029200>
- Meltzer, B., Reichenbach, C. S., Braiman, C., Schiff, N. D., Hudspeth, A. J., & Reichenbach, T. (2015). The steady-state response of the cerebral cortex to the beat of music reflects both the comprehension of music and attention. *Frontiers in Human Neuroscience*, 9(AUGUST), 1–11. <https://doi.org/10.3389/fnhum.2015.00436>
- Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Neuron*, 485(7397). <https://doi.org/10.1038/jid.2014.371>
- Mesgarani, N., David, S. V., Fritz, J. B., & Shamma, S. A. (2009). Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. *Journal of Neurophysiology*, 102(6), 3329–3339. <https://doi.org/10.1152/jn.91128.2008>
- MettingVanRijn, A. C., Kuiper, A. P., Dankers, T. E., & Grimbergen, C. A. (1996, October). Low-cost active electrode improves the resolution in biopotential recordings. In *Proceedings of 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Vol. 1, pp. 101-102)*. IEEE.
- Meyer, J., Oguz, P. G., & Moore, K. S. (2018). Superior fluid cognition in trained musicians. *Psychology of Music*, 48(3), 434–447. <https://doi.org/10.1177/0305735618808089>

- Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, 219(1-2), 36-47. <https://doi.org/10.1016/j.heares.2006.05.004>
- Micheyl, C., & Oxenham, A. J. (2010). Pitch, harmonicity and concurrent sound segregation: Psychoacoustical and neurophysiological findings. *Hearing Research*, 266(1-2), 36-51. <https://doi.org/10.1016/j.heares.2009.09.012>
- Michie, P. T., Bearpark, H. M., Crawford, J. M., & Glue, L. C. (1990). The nature of selective attention effects on auditory event-related potentials. *Biological psychology*, 30(3), 219-250.
- Michie, P. T., Solowij, N., Crawford, J. M., & Glue, L. C. (1993). The effects of between-source discriminability on attended and unattended auditory ERPs. *Psychophysiology*, 30(2), 205-220.
- Middlebrooks, J. C. (2017). Spatial stream segregation. *The auditory system at the cocktail party*, 137-168.
- Middlebrooks, J. C., & Bremen, P. (2013). Spatial stream segregation by auditory cortical neurons. *Journal of Neuroscience*, 33(27), 10986-11001. <https://doi.org/10.1523/JNEUROSCI.1065-13.2013>
- Migirov, L., Kronenberg, J., & Henkin, Y. (2009). Self-reported listening habits and enjoyment of music among adult cochlear implant recipients. *Annals of Otology, Rhinology and Laryngology*, 118(5), 350-355. <https://doi.org/10.1177/000348940911800506>
- Milham, M. P., Banich, M. T., Webb, A., Barad, V., Cohen, N. J., Wszalek, T., & Kramer, A. F. (2001). The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. *Cognitive Brain Research*, 12(3), 467-473. [https://doi.org/10.1016/S0926-6410\(01\)00076-3](https://doi.org/10.1016/S0926-6410(01)00076-3)
- Miller, E., & Cohen, J. D. (2001). AN INTEGRATIVE THEORY OF PREFRONTAL CORTEX FUNCTION. *Annual Review of Neuroscience*, 24, 167-202. <https://doi.org/10.1128/jb.106.1.294-295.1971>
- Miller, G. A., & Heise, G. A. (1950). The trill threshold. *The Journal of the Acoustical Society of America*, 22(5), 637-638.
- Miran, S., Akram, S., Sheikhattar, A., Simon, J. Z., Zhang, T., & Babadi, B. (2018). Real-time tracking of selective auditory attention from M/EEG: A Bayesian filtering approach. *Frontiers in Neuroscience*, 12(MAY), 1-20. <https://doi.org/10.3389/fnins.2018.00262>
- Miran, S., Akram, S., Sheikhattar, A., Simon, J. Z., Zhang, T., & Babadi, B. (2018). Real-Time Decoding of Auditory Attention from EEG via Bayesian Filtering. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2018-July(November)*, 25-28. <https://doi.org/10.1109/EMBC.2018.8512210>
- Miranda, E. R., Magee, W. L., Wilson, J. J., Eaton, J., & Palaniappan, R. (2011). Brain-Computer Music Interfacing (BCMI): From Basic Research to the Real World of Special Needs. *Music & Medicine*, 3(3), 134-140. <https://doi.org/10.1177/1943862111399290>
- Mirkovic, B., Debener, S., Jaeger, M., & De Vos, M. (2015). Decoding the attended speech stream with multi-channel EEG: Implications for online, daily-life applications. *Journal of Neural Engineering*, 12(4). <https://doi.org/10.1088/1741-2560/12/4/046007>
- Mirza, S., Douglas, S. A., Lindsey, P., Hildreth, T., & Hawthorne, M. (2003). Appreciation of music in adult patients with cochlear implants: a patient questionnaire. *Cochlear implants international*, 4(2), 85-95.
- Misurelli, S. M., & Litovsky, R. Y. (2015). Spatial release from masking in children with bilateral cochlear implants and with normal hearing: Effect of target-interferer similarity. *The Journal of the Acoustical Society of America*, 138(1), 319-331. <https://doi.org/10.1121/1.4922777>
- 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>

- Moerel, M., De Martino, F., & Formisano, E. (2014). An anatomical and functional topography of human auditory cortical areas. *Frontiers in Neuroscience*, 8(8 JUL), 1–14. <https://doi.org/10.3389/fnins.2014.00225>
- Molloy, K., Lavie, N., & Chait, M. (2019). Auditory figure-ground segregation is impaired by high visual load. *Journal of Neuroscience*, 39(9), 1699–1708. <https://doi.org/10.1523/JNEUROSCI.2518-18.2018>
- Monahan, C. B., Kendall, R. A., & Carterette, E. C. (1987). The effect of melodic and temporal contour on recognition memory for pitch change. *Perception & Psychophysics*, 41(6), 576–600. <https://doi.org/10.3758/BF03210491>
- Moore, B. C. J. (2003). Speech processing for the hearing-impaired: Successes, failures, and implications for speech mechanisms. *Speech Communication*, 41(1), 81–91. [https://doi.org/10.1016/S0167-6393\(02\)00095-X](https://doi.org/10.1016/S0167-6393(02)00095-X)
- Moore, B. C. J., Glasberg, B. R., & Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *Journal of the Acoustical Society of America*, 80(2), 479–483. <https://doi.org/10.1121/1.394043>
- Moore, B. C. J., & Gockel, H. E. (2012). Properties of auditory stream formation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1591), 919–931. <https://doi.org/10.1098/rstb.2011.0355>
- Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, 11, 56–60.
- Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychological Science*, 22(11), 1425–1433. <https://doi.org/10.1177/0956797611416999>
- Moreno, S., & Bidelman, G. M. (2014). Examining neural plasticity and cognitive benefit through the unique lens of musical training. *Hearing Research*, 308, 84–97. <https://doi.org/10.1016/j.heares.2013.09.012>
- Moreno, S., & Farzan, F. (2015). Music training and inhibitory control: A multidimensional model. *Annals of the New York Academy of Sciences*, 1337(1), 147–152. <https://doi.org/10.1111/nyas.12674>
- Moreno, S., Wodniecka, Z., Tays, W., Alain, C., & Bialystok, E. (2014). Inhibitory control in bilinguals and musicians: Event related potential (ERP) evidence for experience-specific effects. *PLoS ONE*, 9(4), 1–8. <https://doi.org/10.1371/journal.pone.0094169>
- Morey, C. C., & Cowan, N. (2005). When do Visual and Verbal Memories Conflict? The Importance of Working-Memory Load and Retrieval. *Journal of Experimental Psychology: Learning Memory and Cognition*, 31(4), 703–713. <https://doi.org/10.1037/0278-7393.31.4.703.When>
- Morgan, A. L. R., & Brandt, J. F. (1989). An auditory Stroop effect for pitch, loudness, and time. *Brain and Language*, 36(4), 592–603. [https://doi.org/10.1016/0093-934X\(89\)90088-6](https://doi.org/10.1016/0093-934X(89)90088-6)
- Most, S. B., Sorber, A. V., & Cunningham, J. G. (2007). Auditory Stroop reveals implicit gender associations in adults and children. *Journal of Experimental Social Psychology*, 43(2), 287–294. <https://doi.org/10.1016/j.jesp.2006.02.002>
- Mountcastle, V. B. (1997). The columnar organization of the neocortex. *Brain*, 120(4), 701–722. <https://doi.org/10.1093/brain/120.4.701>
- Moussard, A., Bermudez, P., Alain, C., Tays, W., & Moreno, S. (2016). Life-long music practice and executive control in older adults: An event-related potential study. *Brain Research*, 1642, 146–153. <https://doi.org/10.1016/j.brainres.2016.03.028>
- Mrazek, M. D., Smallwood, J., Franklin, M. S., Chin, J. M., Baird, B., & Schooler, J. W. (2012). The role of mind-wandering in measurements of general aptitude. *Journal of experimental psychology: general*, 141(4), 788.

- Mullane, J. C., Corkum, P. V., Klein, R. M., & McLaughlin, E. (2009). Interference control in children with and without ADHD: A systematic review of flanker and simon task performance. *Child Neuropsychology*, 15(4), 321–342. <https://doi.org/10.1080/09297040802348028>
- Munoz, D. P., & Everling, S. (2004). Look away: The anti-saccade task and the voluntary control of eye movement. *Nature Reviews Neuroscience*, 5(3), 218–228. <https://doi.org/10.1038/nrn1345>
- Münste, T. F., Kohlmetz, C., Nager, W., & Altenmüller, E. (2001). Superior auditory spatial tuning in conductors. *Nature*, 409(6820), 580. <https://doi.org/10.1038/35054668>
- Murakami, S., & Okada, Y. (2006). Contributions of principal neocortical neurons to magnetoencephalography and electroencephalography signals. *Journal of Physiology*, 575(3), 925–936. <https://doi.org/10.1113/jphysiol.2006.105379>
- Murphy, S., Spence, C., & Dalton, P. (2017). Auditory perceptual load: A review. *Hearing Research*, 352, 40–48. <https://doi.org/10.1016/j.heares.2017.02.005>
- Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences of the United States of America*, 104(40), 15894–15898. <https://doi.org/10.1073/pnas.0701498104>
- Musacchia, G., Strait, D., & Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hearing Research*, 241(1–2), 34–42. <https://doi.org/10.1016/j.heares.2008.04.013>
- Nagathil, A., Weihs, C., Neumann, K., & Martin, R. (2017). Spectral complexity reduction of music signals based on frequency-domain reduced-rank approximations: An evaluation with cochlear implant listeners. *The Journal of the Acoustical Society of America*, 142(3), 1219–1228.
- Nager, W., Kohlmetz, C., Altenmüller, E., Rodriguez-Fornells, A., & Münste, T. F. (2003). The fate of sounds in conductors' brains: An ERP study. *Cognitive Brain Research*, 17(1), 83–93. [https://doi.org/10.1016/S0926-6410\(03\)00083-1](https://doi.org/10.1016/S0926-6410(03)00083-1)
- Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology*, 53, 53–81. <https://doi.org/10.1146/annurev.psych.53.100901.135131>
- Navon, D. (1984). Resources—a theoretical soup stone? *Psychological Review*, 91(2), 216–234. <https://doi.org/10.1037/0033-295X.91.2.216>
- Neath, I., & Brown, G. D. A. (2012). Arguments against memory trace decay: A SIMPLE account of Baddeley and Scott. *Frontiers in Psychology*, 3(FEB), 2–4. <https://doi.org/10.3389/fpsyg.2012.00035>
- Nee, D. E., Brown, J. W., Askren, M. K., Berman, M. G., Demiralp, E., Krawitz, A., & Jonides, J. (2013). A meta-analysis of executive components of working memory. *Cerebral Cortex*, 23(2), 264–282. <https://doi.org/10.1093/cercor/bhs007>
- Nelson, M. D., Wilson, R. H., & Kornhass, S. (2003). Performance of musicians and non musicians on dichotic chords, dichotic CVs, and dichotic digits. *Journal of the American Academy of Audiology*, 14(10), 536–549. <https://doi.org/10.3766/jaaa.14.10.2>
- Nelson, P. B., & Jin, S.-H. (2004). Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners. *The Journal of the Acoustical Society of America*, 115(5), 2286–2294. <https://doi.org/10.1121/1.1703538>
- Nelson, P. B., Jin, S.-H., Carney, A. E., & Nelson, D. A. (2003). Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners. *The Journal of the Acoustical Society of America*, 113(2), 961–968. <https://doi.org/10.1121/1.1531983>
- Newman, R. S., & Evers, S. (2007). The effect of talker familiarity on stream segregation. *Journal of Phonetics*, 35(1), 85–103. <https://doi.org/10.1016/j.wocn.2005.10.004>
- Nobre, A. C., & Stokes, M. G. (2011). Attention and short-term memory: crossroads.

- Noonan, M. A. P., Crittenden, B. M., Jensen, O., & Stokes, M. G. (2018). Selective inhibition of distracting input. *Behavioural Brain Research*, 355(October 2017), 36–47. <https://doi.org/10.1016/j.bbr.2017.10.010>
- Nöstl, A., Marsh, J. E., & Sörqvist, P. (2012). Expectations Modulate the Magnitude of Attentional Capture by Auditory Events. *PLoS ONE*, 7(11). <https://doi.org/10.1371/journal.pone.0048569>
- 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. *Journal of Neuroscience*, 29(49), 15564–15574. <https://doi.org/10.1523/JNEUROSCI.3065-09.2009>
- Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception and Psychophysics*, 60(3), 355–376. <https://doi.org/10.3758/BF03206860>
- Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, 5(1), 42–46. <https://doi.org/10.1111/j.1467-9280.1994.tb00612.x>
- 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>
- O’Sullivan, J., Chen, Z., Sheth, S. A., McKhann, G., Mehta, A. D., & Mesgarani, N. (2017). Neural decoding of attentional selection in multi-speaker environments without access to separated sources. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 1644–1647. <https://doi.org/10.1109/EMBC.2017.8037155>
- Oberauer, K. (2019). Working memory and attention - A conceptual analysis and review. *Journal of Cognition*, 2(1), 1–23. <https://doi.org/10.5334/joc.58>
- Oberauer, K., & Lin, H. Y. (2017). An interference model of visual working memory. *Psychological Review*, 124(1), 21–59. <https://doi.org/10.1037/rev0000044>
- Oberauer, K. (2005). Control of the contents of working memory--a comparison of two paradigms and two age groups. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(4), 714.
- Oberauer, K., Süß, H., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity – Facets of a cognitive ability construct . *Personality and Individual Dif. Personality and Individual Differences*, 29, 2000–2000. www.elsevier.com/locate/paid
- Olazabal, U., & Moore, J. K. (1984). Acustico-motor pathways in rats demonstrated by retrograde HRP injections. In *Soc. Neurosci. Abstr.* (Vol. 10, p. 351).
- Ollman, R. T. (1973). Simple reactions with random countermanding of the “go” signal. *Attention and performance IV*, 571-581.
- Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, 25(1), 46–59. <https://doi.org/10.1002/hbm.20131>
- Oxenham, A. J., Fligor, B. J., Mason, C. R., & Kidd, G. (2003). Informational masking and musical training. *The Journal of the Acoustical Society of America*, 114(3), 1543–1549. <https://doi.org/10.1121/1.1598197>
- Pallesen, K. J., Brattico, E., Bailey, C. J., Korvenoja, A., Koivisto, J., Gjedde, A., & Carlson, S. (2010). Cognitive control in auditory working memory is enhanced in musicians. *PLoS ONE*, 5(6). <https://doi.org/10.1371/journal.pone.0011120>
- Panek, P. E., & Rush, M. C. (1981). Simultaneous examination of age related differences in the ability to maintain and reorient auditory selective attention. *Experimental aging research*, 7(4), 405-416.
- Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature*, 392(6678), 811–814. <https://doi.org/10.1038/33918>

- Parasuraman, R. (1998). "The attentive brain: issues and prospects," in *The Attentive Brain*, ed R. Parasuraman (Cambridge, MA: The MIT Press), 3–15.
- Parbery-Clark, A., Marmel, F., Bair, J., & Kraus, N. (2011). What subcortical-cortical relationships tell us about processing speech in noise. *European Journal of Neuroscience*, 33(3), 549–557. <https://doi.org/10.1111/j.1460-9568.2010.07546.x>
- Parbery-Clark, A., Skoe, E., & Kraus, N. (2009b). Musical experience limits the degradative effects of background noise on the neural processing of sound. *Journal of Neuroscience*, 29(45), 14100–14107. <https://doi.org/10.1523/JNEUROSCI.3256-09.2009>
- Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009a). Musician enhancement for speech-in-noise. *Ear and Hearing*, 30(6), 653–661. <https://doi.org/10.1097/AUD.0b013e3181b412e9>
- Paré, M., & Hanes, D. P. (2003). Controlled movement processing: Superior colliculus activity associated with countermanded saccades. *Journal of Neuroscience*, 23(16), 6480–6489. <https://doi.org/10.1523/jneurosci.23-16-06480.2003>
- Parmentier, F. B. R., Elsley, J. V., Andrés, P., & Barceló, F. (2011). Why are auditory novels distracting? Contrasting the roles of novelty, violation of expectation and stimulus change. *Cognition*, 119(3), 374–380. <https://doi.org/10.1016/j.cognition.2011.02.001>
- Parmentier, F. B. R., Pacheco-Unguetti, A. P., & Valero, S. (2018). Food words distract the hungry: Evidence of involuntary semantic processing of task-irrelevant but biologically-relevant unexpected auditory words. *PLoS ONE*, 13(1), 1–17. <https://doi.org/10.1371/journal.pone.0190644>
- Pascal, M. (2020). Analyse de la composition de l' espace dans une œuvre acousmatique immersive de Jean Marc Duchenne. *Hal Archives-Ouvertes*, 1–8.
- Pashler, H. (1998). *The psychology of attention*. Cambridge, MA: MIT Press.
- Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., Knight, R. T., & Chang, E. F. (2012). Reconstructing speech from human auditory cortex. *PLoS Biology*, 10(1). <https://doi.org/10.1371/journal.pbio.1001251>
- Patou, F., Py, H., Backus, B., & Riis, S. (2019). Oticon Medical ' s ecosystem of lab- and field investigational platforms for behavioural and cognitive research in CI Peripheral neural interfacing Behavior. October, 2019.
- Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, 23(6), 1378–1387. <https://doi.org/10.1093/cercor/bhs118>
- Penhune, V. B., Zatorre, R. J., MacDonald, J. D., & Evans, A. C. (1996). Interhemispheric anatomical differences in human primary auditory cortex: Probabilistic mapping and volume measurement from magnetic resonance scans. *Cerebral Cortex*, 6(5), 661–672. <https://doi.org/10.1093/cercor/6.5.661>
- Peretz, I., & Zatorre, R. J. (2005). Brain organization for music processing. *Annual Review of Psychology*, 56, 89–114. <https://doi.org/10.1146/annurev.psych.56.091103.070225>
- Petersen, S., & Posner, M. (2012). The Attention System of the Human Brain: 20 Years After. *Annual Review of Neuroscience*, 21(35), 73–89. <https://doi.org/10.1146/annurev-neuro-062111-150525>
- Pickles, J. O. (2015). Auditory pathways: Anatomy and physiology. In *Handbook of Clinical Neurology* (1st ed., Vol. 129). Elsevier B.V. <https://doi.org/10.1016/B978-0-444-62630-1.00001-9>
- Picton, T. W., Stuss, D. T., Alexander, M. P., Shallice, T., Binns, M. A., & Gillingham, S. (2007). Effects of focal frontal lesions on response inhibition. *Cerebral Cortex*, 17(4), 826–838. <https://doi.org/10.1093/cercor/bhk031>
- Pinto, Y., van der Leij, A. R., Sligte, I. G., Lamme, V. A. F., & Scholte, H. S. (2013). Bottom-up and top-down attention are independent. *Journal of Vision*, 13(3), 16. <https://doi.org/10.1167/13.3.16>

- Plakke, B., Hwang, J., Diltz, M., & Romanski, L. M. (2012). Inactivation of ventral prefrontal cortex impairs audiovisual working memory. *Society of Neuroscience New Orleans*, 87810.
- Plomp, R., & Mimpen, A. M. (1981). Effect of the orientation of the speaker's head and the azimuth of a noise source on the speech-reception threshold for sentences. *Acta Acustica United with Acustica*, 48(5), 325-328.
- Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as "asymmetric sampling in time." *Speech Communication*, 41(1), 245-255. [https://doi.org/10.1016/S0167-6393\(02\)00107-3](https://doi.org/10.1016/S0167-6393(02)00107-3)
- Polich, J. (1986). Attention, probability, and task demands as determinants of P300 latency from auditory stimuli. *Electroencephalography and clinical neurophysiology*, 63(3), 251-259.
- Polich, J. (2007). Updating P300: An Integrative Theory of P3a and P3b. *Clinical Neurophysiology*, 118(10), 2128-2148. <https://doi.org/10.1016/j.clinph.2007.04.019>. Updating
- Pons, J., Janer, J., Rode, T., & Nogueira, W. (2016). Remixing music using source separation algorithms to improve the musical experience of cochlear implant users. *The Journal of the Acoustical Society of America*, 140(6), 4338-4349. <https://doi.org/10.1121/1.4971424>
- Ponti, M., Bélanger, S., Grimes, R., Heard, J., Johnson, M., Moreau, E., Norris, M., Shaw, A., Stanwick, R., Van Lankveld, J., & Williams, R. (2017). Screen time and young children: Promoting health and development in a digital world. *Paediatrics and Child Health (Canada)*, 22(8), 461-477. <https://doi.org/10.1093/pch/pxx123>
- Posner, M. I., & Petersen, S. E. (1990). The Attention System of the Human Brain. *Annual Review of Neuroscience*, 13(1), 25-42.
- Posner, M. I. (1980). Orienting or attention. *Quarterly Journal of Experimental Psychology*, 32(1), 3-25.
- Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition* (pp. 55-85). Hillsdale, NJ: Erlbaum
- Posner, M. I., & Fan, J. (2008). Attention as an organ system. *Topics in integrative neuroscience*, 31.
- Pouget, P., Murthy, A., & Stuphorn, V. (2017). Cortical control and performance monitoring of interrupting and redirecting movements. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1718). <https://doi.org/10.1098/rstb.2016.0201>
- Power, A. J., Lalor, E. C., & Reilly, R. B. (2011). Endogenous auditory spatial attention modulates obligatory sensory activity in auditory cortex. *Cerebral Cortex*, 21(6), 1223-1230. <https://doi.org/10.1093/cercor/bhq233>
- Pressnitzer, D. (2011). Auditory scene analysis: the sweet music of ambiguity. *Frontiers in Human Neuroscience*, 5(December), 1-11. <https://doi.org/10.3389/fnhum.2011.00158>
- Pressnitzer, D., & Hupé, J. M. (2006). Temporal Dynamics of Auditory and Visual Bistability Reveal Common Principles of Perceptual Organization. *Current Biology*, 16(13), 1351-1357. <https://doi.org/10.1016/j.cub.2006.05.054>
- PrevotEAU, C., Chen, S. Y., & Lalwani, A. K. (2018). Music enjoyment with cochlear implantation. *Auris Nasus Larynx*, 45(5), 895-902. <https://doi.org/10.1016/j.anl.2017.11.008>
- Puschmann, S., Baillet, S., & Zatorre, R. J. (2019). Musicians at the Cocktail Party: Neural Substrates of Musical Training during Selective Listening in Multispeaker Situations. *Cerebral Cortex*, 29(8), 3253-3265. <https://doi.org/10.1093/cercor/bhy193>
- Puvvada, K. C., & Simon, J. Z. (2017). Cortical representations of speech in a multitalker auditory scene. *Journal of Neuroscience*, 37(38), 9189-9196. <https://doi.org/10.1523/JNEUROSCI.0938-17.2017>

- Quak, M., London, R. E., & Talsma, D. (2015). A multisensory perspective of working memory. *Frontiers in Human Neuroscience*, 9(APR), 1–11. <https://doi.org/10.3389/fnhum.2015.00197>
- Quesnel, A. M., Ishai, R., & McKenna, M. J. (2018). Otosclerosis: temporal bone pathology. *Otolaryngologic Clinics of North America*, 51(2), 291–303.
- Ragert, M., Fairhurst, M. T., & Keller, P. E. (2014). Segregation and integration of auditory streams when listening to multi-part music. *PLoS ONE*, 9(1). <https://doi.org/10.1371/journal.pone.0084085>
- Ramakrishnan, A., Sureshbabu, R., & Murthy, A. (2012). Understanding how the brain changes its mind: Microstimulation in the macaque frontal eye field reveals how saccade plans are changed. *Journal of Neuroscience*, 32(13), 4457–4472. <https://doi.org/10.1523/JNEUROSCI.3668-11.2012>
- Rammsayer, T., & Altenmüller, E. (2006). TEMPORAL INFORMATION PROCESSING IN MUSICIANS AND NON MUSICIANS. *Music Perception*, 24(1), 37–48.
- Rasch, R. A. (1978). The perception of simultaneous notes such as in polyphonic music. *Acta Acustica united with Acustica*, 40(1), 21–33.
- Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, 28(3), 164–171. <https://doi.org/10.1027/1015-5759/a000123>
- Repp, B. H., & Doggett, R. (2007). Tapping to a very slow beat: a comparison of musicians and non musicians. *Music Perception*, 24(4), 367–376.
- Repp, B. H. (2010). Sensorimotor synchronization and perception of timing: effects of music training and task experience. *Human movement science*, 29(2), 200–213.
- Remijn, G. B., & Kojima, H. (2010). Active versus passive listening to auditory streaming stimuli: a near-infrared spectroscopy study. *Journal of Biomedical Optics*, 15(3), 037006. <https://doi.org/10.1117/1.3431104>
- Rimmele, J., Schröger, E., & Bendixen, A. (2012). Age-related changes in the use of regular patterns for auditory scene analysis. *Hearing research*, 289(1-2), 98–107.
- Roberts, B., Glasberg, B. R., & Moore, B. C. J. (2002). Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *The Journal of the Acoustical Society of America*, 112(5), 2074–2085. <https://doi.org/10.1121/1.1508784>
- Roberts, K. L., & Hall, D. A. (2008). Examining a supramodal network for conflict processing: A systematic review and novel functional magnetic resonance imaging data for related visual and auditory stroop tasks. *Journal of Cognitive Neuroscience*, 20(6), 1063–1078. <https://doi.org/10.1162/jocn.2008.20074>
- Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). “Oops!” Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Peabody Journal of Education*, 35(6), 747–758. <https://doi.org/10.1080/01619565609536671>
- Roden, I., Könen, T., Bongard, S., Frankenberg, E., Friedrich, E. K., & Kreutz, G. (2014). Effects of music training on attention, processing speed and cognitive music abilities-findings from a longitudinal study. *Applied Cognitive Psychology*, 28(4), 545–557. <https://doi.org/10.1002/acp.3034>
- Román-Caballero, R., Martín-Arévalo, E., & Lupiáñez, J. (2021). Attentional networks functioning and vigilance in expert musicians and non-musicians. *Psychological Research*, 85(3), 1121–1135. <https://doi.org/10.1007/s00426-020-01323-2>

- Roman, S., Rochette, F., Triglia, J. M., Schön, D., & Bigand, E. (2016). Auditory training improves auditory performance in cochlear implanted children. *Hearing Research*, 337, 89–95. <https://doi.org/10.1016/j.heares.2016.05.003>
- Rosenboom, D. (1977). Biofeedback and the arts: Results of early experiments. *Journal of Aesthetics and Art Criticism*, 35(3).
- Ruby, F. J. M., Smallwood, J., Engen, H., & Singer, T. (2013). How Self-Generated Thought Shapes Mood-The Relation between Mind-Wandering and Mood Depends on the Socio-Temporal Content of Thoughts. *PLoS ONE*, 8(10). <https://doi.org/10.1371/journal.pone.0077554>
- Ruby, F. J. M., Smallwood, J., Sackur, J., & Singer, T. (2013). Is self-generated thought a means of social problem solving? *Frontiers in Psychology*, 4(DEC), 1–10. <https://doi.org/10.3389/fpsyg.2013.00962>
- Ruggles, D. R., Freyman, R. L., & Oxenham, A. J. (2014). Influence of musical training on understanding voiced and whispered speech in noise. *PLoS ONE*, 9(1). <https://doi.org/10.1371/journal.pone.0086980>
- Rüsseler, J., Hennighausen, E., Münte, T. F., & Rösler, F. (2003). Differences in incidental and intentional learning of sensorimotor sequences as revealed by event-related brain potentials. *Cognitive Brain Research*, 15(2), 116–126. [https://doi.org/10.1016/S0926-6410\(02\)00145-3](https://doi.org/10.1016/S0926-6410(02)00145-3)
- Sach, A. J., & Bailey, P. J. (2004). Some characteristics of auditory spatial attention revealed using rhythmic masking release. *Perception & Psychophysics*, 66(8), 1379–1387.
- Sachs, M., Kaplan, J., Der Sarkissian, A., & Habibi, A. (2017). Increased engagement of the cognitive control network associated with music training in children during an fMRI Stroop task. *PLoS ONE*, 12(10), 1–29. <https://doi.org/10.1371/journal.pone.0187254>
- Saults, J. S., & Cowan, N. (2007). A Central Capacity Limit to the Simultaneous Storage of Visual and Auditory Arrays in Working Memory. *Journal of Experimental Psychology: General*, 136(4), 663–684. <https://doi.org/10.1037/0096-3445.136.4.663.A>
- Saupe, K., Koelsch, S., & Rübsem, R. (2010). Spatial selective attention in a complex auditory environment such as polyphonic music. *The Journal of the Acoustical Society of America*, 127(1), 472–480. <https://doi.org/10.1121/1.3271422>
- Savage, R., Lavers, N., & Pillay, V. (2007). Working memory and reading difficulties: What we know and what we don't know about the relationship. *Educational Psychology Review*, 19(2), 185–221. <https://doi.org/10.1007/s10648-006-9024-1>
- Schaefer, R. S., Farquhar, J., Blokland, Y., Sadakata, M., & Desain, P. (2011). Name that tune: Decoding music from the listening brain. *NeuroImage*, 56(2), 843–849. <https://doi.org/10.1016/j.neuroimage.2010.05.084>
- Schäfer, P. J., Corona-Strauss, F. I., Hannemann, R., Hillyard, S. A., & Strauss, D. J. (2018). Testing the Limits of the Stimulus Reconstruction Approach: Auditory Attention Decoding in a Four-Speaker Free Field Environment. *Trends in Hearing*, 22, 1–12. <https://doi.org/10.1177/2331216518816600>
- Schall, J. D., Palmeri, T. J., & Logan, G. D. (2017). Models of inhibitory control. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1718). <https://doi.org/10.1098/rstb.2016.0193>
- Schellenberg, E. G. (2011). Examining the association between music lessons and intelligence. *British Journal of Psychology*, 102(3), 283–302. <https://doi.org/10.1111/j.2044-8295.2010.02000.x>
- Schellenberg, E. G., & Peretz, I. (2008). Music, language and cognition: unresolved issues.
- Schlaug, G., Jäncke, L., Huang, Y., Staiger, J. F., & Steinmetz, H. (1995). Increased corpus callosum size in musicians. *Neuropsychologia*, 33(8), 1047–1055. [https://doi.org/10.1016/0028-3932\(95\)00045-5](https://doi.org/10.1016/0028-3932(95)00045-5)
- Schleich, P., Nopp, P., & D'haese, P. (2004). Head shadow, squelch, and summation effects in bilateral users of the MED-EL COMBI 40/40+ cochlear implant. *Ear and hearing*, 25(3), 197–204.

- Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience*, 5(7), 688–694. <https://doi.org/10.1038/nn871>
- Schön, D., & François, C. (2011). Musical expertise and statistical learning of musical and linguistic structures. *Frontiers in Psychology*, 2(JUL), 1–9. <https://doi.org/10.3389/fpsyg.2011.00167>
- Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, 41(3), 341–349. <https://doi.org/10.1111/1469-8986.00172.x>
- Schooler, J. W., Reichle, E. D., & Halpern, D. V. (2004). Zoning out while reading: Evidence for dissociations between experience and metaconsciousness. MIT press.
- Schröger, E., Giard, M. H., & Wolff, C. (2000). Auditory distraction: Event-related potential and behavioral indices. *Clinical Neurophysiology*, 111(8), 1450–1460. [https://doi.org/10.1016/S1388-2457\(00\)00337-0](https://doi.org/10.1016/S1388-2457(00)00337-0)
- Schulze, K., Mueller, K., & Koelsch, S. (2011a). Neural correlates of strategy use during auditory working memory in musicians and non-musicians. *European Journal of Neuroscience*, 33(1), 189–196. <https://doi.org/10.1111/j.1460-9568.2010.07470.x>
- Schulze, K., Jay Dowling, W., & Tillmann, B. (2012). Working Memory for Tonal and Atonal Sequences during a Forward and a Backward Recognition Task. *Music Perception: An Interdisciplinary Journal*, 29(3), 255–267. <https://doi.org/10.1111/j.1835-9310.1982.tb01239.x>
- Schulze, K., & Koelsch, S. (2012). Working memory for speech and music. *Annals of the New York Academy of Sciences*, 1252(1), 229–236. <https://doi.org/10.1111/j.1749-6632.2012.06447.x>
- Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., & Koelsch, S. (2011b). Neuroarchitecture of verbal and tonal working memory in non musicians and musicians. *Human Brain Mapping*, 32(5), 771–783. <https://doi.org/10.1002/hbm.21060>
- Schwartz, J. L., Grimault, N., Hupé, J. M., Moore, B. C. J., & Pressnitzer, D. (2012). Multistability in perception: Binding sensory modalities, an overview. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1591), 896–905. <https://doi.org/10.1098/rstb.2011.0254>
- Scott, S. K., Catrin Blank, C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, 123(12), 2400–2406. <https://doi.org/10.1093/brain/123.12.2400>
- Seidman, M. D., Ahmad, N., & Bai, U. (2002). Molecular mechanisms of age-related hearing loss. *Ageing Research Reviews*, 1(3), 331–343. [https://doi.org/10.1016/S1568-1637\(02\)00004-1](https://doi.org/10.1016/S1568-1637(02)00004-1)
- Seli, P., Cheyne, J. A., Barton, K. R., & Smilek, D. (2012). Consistency of sustained attention across modalities: Comparing visual and auditory versions of the SART. *Canadian Journal of Experimental Psychology*, 66(1), 44–50. <https://doi.org/10.1037/a0025111>
- Seli, P., Jonker, T. R., Cheyne, J. A., & Smilek, D. (2013). Enhancing SART validity by statistically controlling speed-accuracy trade-offs. *Frontiers in Psychology*, 4(MAY), 1–8. <https://doi.org/10.3389/fpsyg.2013.00265>
- Seli, P., Kane, M. J., Smallwood, J., Schacter, D. L., Maillet, D., Schooler, J. W., & Smilek, D. (2018). Mind-Wandering as a Natural Kind: A Family-Resemblances View. *Trends in Cognitive Sciences*, 22(6), 479–490. <https://doi.org/10.1016/j.tics.2018.03.010>
- Seli, P., Risko, E. F., & Smilek, D. (2016). On the Necessity of Distinguishing Between Unintentional and Intentional Mind Wandering. *Psychological Science*, 27(5), 685–691. <https://doi.org/10.1177/0956797616634068>
- Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *Journal of Neuroscience*, 23(13), 5545–5552. <https://doi.org/10.1523/jneurosci.23-13-05545.2003>

- Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*, 34(3), 114–123. <https://doi.org/10.1016/j.tins.2010.11.002>. Temporal
- Shamma, S. A., & Micheyl, C. (2010). Behind the scenes of auditory perception. *Current Opinion in Neurobiology*, 20(3), 361–366. <https://doi.org/10.1016/j.conb.2010.03.009>
- Shannon, R. V, Zeng, F., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech Recognition with Primarily Temporal Cues. *Science*, 270(5234), 303–304.
- Sharma, V. V., Thaut, M., Russo, F., & Alain, C. (2019). Absolute Pitch and Musical Expertise Modulate Neuro-Electric and Behavioral Responses in an Auditory Stroop Paradigm. *Frontiers in Neuroscience*, 13(September), 1–14. <https://doi.org/10.3389/fnins.2019.00932>
- Sheft, S. (2007). Envelope Processing and Sound-Source Perception. *Auditory Perception of Sound Sources*, 233–280. https://doi.org/10.1007/978-0-387-71305-2_9
- Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention Barbara. *Trends in Cognitive Sciences*, 12(5), 182–186. <https://doi.org/10.1016/j.tics.2008.02.003>. Object-based
- Shinn-Cunningham, B. G., Lee, A. K. C., & Oxenham, A. J. (2007). A sound element gets lost in perceptual competition. *Proceedings of the National Academy of Sciences of the United States of America*, 104(29), 12223–12227. <https://doi.org/10.1073/pnas.0704641104>
- Shinn-Cunningham, B., Best, V., Kingstone, A., Fawcett, J. M., & Risko, E. F. (2015). Auditory selective attention. *The handbook of attention*, 99.
- Shinn-Cunningham, B., Best, V., & Lee, A. K. (2017). Auditory object formation and selection. In *The auditory system at the cocktail party* (pp. 7-40). Springer, Cham.
- Simmonds, D., Pekaram, J., & Stewart, H. M. (2008). Meta-analysis of Go/No-go tasks demonstrating that fMRI activation associated with response inhibition is task-dependent. *Neuropsychologia*, 4(164), 224–232.
- Simon, A., Østergaard, J., Bech, S., & Loquet, G. (2022). Optimal time lags for linear cortical auditory attention detection : differences between speech and music listening. 19th International Symposium on Hearing 19-24 June 2022, Lyon, France, June. <https://doi.org/10.5281/zenodo.6576990>
- Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of experimental psychology*, 81(1), 174.
- Simon, J. R., & Berbaum, K. (1990). Effect of conflicting cues on information processing: the ‘Stroop effect’ vs. the ‘Simon effect’. *Acta psychologica*, 73(2), 159-170.
- Simon, J. R., & Craft, J. L. (1970). Effects of an irrelevant auditory stimulus on visual choice reaction time. *Journal of experimental psychology*, 86(2), 272.
- Singh, P. G. (1987). Perceptual organization of complex-tone sequences: A tradeoff between pitch and timbre?. *The Journal of the Acoustical Society of America*, 82(3), 886-899.
- Singh, P. G., & Bregman, A. S. (1997). The influence of different timbre attributes on the perceptual segregation of complex-tone sequences. *The Journal of the acoustical society of America*, 102(4), 1943-1952.
- Sirevaag, E. J., Kramer, A. F., Coles, M. G., & Donchin, E. (1989). Resource reciprocity: An event-related brain potentials analysis. *Acta psychologica*, 70(1), 77-97.
- Skaansar, J. F., Laeng, B., & Danielsen, A. (2019). Microtiming and mental effort: Onset asynchronies in musical rhythm modulate pupil size. *Music Perception*, 37(2), 111–133. <https://doi.org/10.1525/MP.2019.37.2.111>
- Skuk, V. G., & Schweinberger, S. R. (2013). Gender differences in familiar voice identification. *Hearing Research*, 296, 131–140. <https://doi.org/10.1016/j.heares.2012.11.004>

- Sladen, D. P., & Zappler, A. (2015). Older and Younger Adult Cochlear Implant Users: Speech Recognition in Quiet and Noise, Quality of Life, and Music Perception. *American Journal of Audiology*, 24, 31–39. https://doi.org/10.1044/2014_AJA-13-0066
- Smallwood, J. (2013). Distinguishing how from why the mind wanders: A process-occurrence framework for self-generated mental activity. *Psychological Bulletin*, 139(3), 519–535. <https://doi.org/10.1037/a0030010>
- Smallwood, J. M., Baracaia, S. F., Lowe, M., & Obonsawin, M. (2003). Task unrelated thought whilst encoding information. *Consciousness and Cognition*, 12(3), 452–484. [https://doi.org/10.1016/S1053-8100\(03\)00018-7](https://doi.org/10.1016/S1053-8100(03)00018-7)
- Smallwood, J., Brown, K. S., Baird, B., Mrazek, M. D., Franklin, M. S., & Schooler, J. W. (2012). Insulation for daydreams: A role for tonic norepinephrine in the facilitation of internally guided thought. *PLoS ONE*, 7(4), 1–5. <https://doi.org/10.1371/journal.pone.0033706>
- Smallwood, J., Brown, K. S., Tipper, C., Giesbrecht, B., Franklin, M. S., Mrazek, M. D., Carlson, J. M., & Schooler, J. W. (2011). Pupillometric evidence for the decoupling of attention from perceptual input during offline thought. *PLoS ONE*, 6(3). <https://doi.org/10.1371/journal.pone.0018298>
- Smallwood, J., McSpadden, M., & Schooler, J. W. (2008). When attention matters: The curious incident of the wandering mind. *Memory and Cognition*, 36(6), 1144–1150. <https://doi.org/10.3758/MC.36.6.1144>
- Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, 132(6), 946–958. <https://doi.org/10.1037/0033-2909.132.6.946>
- Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: Empirically navigating the stream of consciousness. *Annual Review of Psychology*, 66, 487–518. <https://doi.org/10.1146/annurev-psych-010814-015331>
- Smith, J., Hausfeld, S., Power, R. P., & Gorta, A. (1982). Ambiguous musical figures and auditory streaming. *Perception & Psychophysics*, 32(5), 454–464. <https://doi.org/10.3758/BF03202776>
- Smith, P. H., Joris, P. X., Banks, M. I., & Yin, T. C. (1993). Responses of cochlear nucleus cells and projections of their axons. In *The Mammalian Cochlear Nuclei* (pp. 349–360). Springer, Boston, MA.
- Snoeckx, R. L., Huygen, P. L. M., & Feldmann, D. (2005). GJB2 Mutations and Degree of Hearing Loss: A Multicenter Study. *The American Journal of Otology*, 26(12), 945–957.
- Snyder, J. S., & Alain, C. (2007). Toward a Neurophysiological Theory of Auditory Stream Segregation. *Psychological Bulletin*, 133(5), 780–799. <https://doi.org/10.1037/0033-2909.133.5.780>
- Snyder, J. S., Alain, C., & Picton W., T. (2006). Effects of Attention on Neuroelectric Correlates of Auditory Stream Segregation. *Journal of Cognitive Neuroscience*, 18(1), 1–13.
- Snyder, J. S., Gregg, M. K., Weintraub, D. M., & Alain, C. (2012). Attention, awareness, and the perception of auditory scenes. *Frontiers in Psychology*, 3(FEB), 1–17. <https://doi.org/10.3389/fpsyg.2012.00015>
- Sogebi, O. A., Oyewole, E. A., Manifah, T. O., & Ogunbanwo, O. (2017). Hearing dynamics in patients with traumatic tympanic membrane perforation. *Journal of the West African College of Surgeons*, 7(2), 15.
- Sohoglu, E., & Chait, M. (2016). Detecting and representing predictable structure during auditory scene analysis. *ELife*, 5(Se), 1–17. <https://doi.org/10.7554/eLife.19113>
- Sommers, M. S., & Danielson, S. M. (1999). Inhibitory processes and spoken word recognition in young and older adults: The interaction of lexical competition and semantic context. *Psychology and Aging*, 14(3), 458–472. <https://doi.org/10.1037/0882-7974.14.3.458>
- Soni, S., & Tata, M. S. (2021). Brain electrical dynamics in speech segmentation depends upon prior experience with the language. *Brain and Language*, 219(May), 104967. <https://doi.org/10.1016/j.bandl.2021.104967>

- Sörqvist, P., & Rönnerberg, J. (2012). Episodic long-term memory of spoken discourse masked by speech: What is the role for working memory capacity? *Journal of Speech, Language, and Hearing Research*, 55(1), 210–218. [https://doi.org/10.1044/1092-4388\(2011/10-0353\)](https://doi.org/10.1044/1092-4388(2011/10-0353))
- Southwell, R., Baumann, A., Gal, C., Barascud, N., Friston, K., & Chait, M. (2017). Is predictability salient? A study of attentional capture by auditory patterns. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1714). <https://doi.org/10.1098/rstb.2016.0105>
- Souza, P. E., Arehart, K. H., Shen, J., Anderson, M., & Kates, J. M. (2015). Working memory and intelligibility of hearing-aid processed speech. *Frontiers in Psychology*, 6(MAY), 1–14. <https://doi.org/10.3389/fpsyg.2015.00526>
- Souza, P., Gehani, N., Wright, R., & McCloy, D. (2013). The advantage of knowing the talker. *Journal of the American Academy of Audiology*, 24(8), 689–700. <https://doi.org/10.3766/jaaa.24.8.6>
- Spagna, A., Mackie, M. A., & Fan, J. (2015). Supramodal executive control of attention. *Frontiers in Psychology*, 6(FEB), 1–14. <https://doi.org/10.3389/fpsyg.2015.00065>
- Spelke, E., Hirst, W., & Neisser, U. (1976). Skills of divided attention. *Cognition*, 4, 215±230.
- Spitzer, M. W., & Semple, M. N. (1995). Neurons sensitive to interaural phase disparity in gerbil superior olive: diverse monaural and temporal response properties. *Journal of neurophysiology*, 73(4), 1668–1690.
- St Clair-Thompson, H. L. (2010). Backwards digit recall: A measure of short-term memory or working memory? *European Journal of Cognitive Psychology*, 22(2), 286–296. <https://doi.org/10.1080/09541440902771299>
- Strait, D. L., & Kraus, N. (2011). Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. *Frontiers in Psychology*, 2(JUN), 1–10. <https://doi.org/10.3389/fpsyg.2011.00113>
- Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research*, 261(1–2), 22–29. <https://doi.org/10.1016/j.heares.2009.12.021>
- Strait, D. L., Slater, J., Abecassis, V., & Kraus, N. (2014). Cortical response variability as a developmental index of selective auditory attention. *Developmental Science*, 17(2), 175–186. <https://doi.org/10.1111/desc.12107>
- Strait, D. L., Slater, J., O'Connell, S., & Kraus, N. (2015). Music training relates to the development of neural mechanisms of selective auditory attention. *Developmental Cognitive Neuroscience*, 12, 94–104. <https://doi.org/10.1016/j.dcn.2015.01.001>
- Strauß, A., Wöstmann, M., & Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Frontiers in Human Neuroscience*, 8(MAY), 1–7. <https://doi.org/10.3389/fnhum.2014.00350>
- Striem-Amit, E., Hertz, U., & Amedi, A. (2011). Extensive cochleotopic mapping of human auditory cortical fields obtained with phase-encoding fMRI. *PLoS ONE*, 6(3). <https://doi.org/10.1371/journal.pone.0017832>
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, 18(6), 643.
- Strouwen, C., Centre, N., Mu, L., Sciences, R., Bloem, B. R., & Nieuwboer, A. (2016). Test-Retest Reliability of Dual-Task Outcome Measures in People With Parkinson Disease. *Physical Therapy*, 96(8), 1276–1286.
- Sturm, I., Dähne, S., Blankertz, B., & Curio, G. (2015). Multi-variate EEG analysis as a novel tool to examine brain responses to naturalistic music stimuli. *PLoS ONE*, 10(10), 1–30. <https://doi.org/10.1371/journal.pone.0141281>
- Sturm, W., & Willmes, K. (2001). On the functional neuroanatomy of intrinsic and phasic alertness. *NeuroImage*, 14(1 II). <https://doi.org/10.1006/nimg.2001.0839>

- Stuss DT, Shallice T, Alexander MP, P. T. (1995). A Multidisciplinary Approach to Anterior Attentional Functions. *ANNALS the New York Academy of Sciences*, 15(769), 191–211.
- Suárez, L., Elangovan, S., & Au, A. (2016). Cross-sectional study on the relationship between music training and working memory in adults. *Australian Journal of Psychology*, 68(1), 38–46. <https://doi.org/10.1111/ajpy.12087>
- Sussman, E. S. (2017). Auditory Scene Analysis: An Attention Perspective. *Journal of Speech, Language, and Hearing Research*. <https://doi.org/10.1044/2017>
- Sussman, E. S., Horvath, J., Winkler, I., & Orr, M. (2007). The role of attention in the formation of auditory streams. *Perception & Psychophysics*, 69(1), 136–152. <https://doi.org/10.1080/21674086.1978.11926859>
- Sutherland, D. P., Masterton, R. B., & Glendenning, K. K. (1998). Role of acoustic striae in hearing: reflexive responses to elevated sound-sources. *Behavioural brain research*, 97(1-2), 1-12.
- Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Kidd, G., & Patel, A. D. (2015). Musical training, individual differences and the cocktail party problem. *Scientific Reports*, 5, 1–11. <https://doi.org/10.1038/srep11628>
- Taheri, B. A., Knight, R. T., & Smith, R. L. (1994). A dry electrode for EEG recording. *Electroencephalography and clinical neurophysiology*, 90(5), 376-383.
- Talamini, F., Altoè, G., Carretti, B., & Grassi, M. (2017). Musicians have better memory than non musicians: A meta-analysis. *PLoS ONE*, 12(10), 1–21. <https://doi.org/10.1371/journal.pone.0186773>
- Talamini, F., Blain, S., Ginzburg, J., Houix, O., Bouchet, P., Grassi, M., Tillmann, B., & Caclin, A. (2022). Auditory and visual short-term memory: influence of material type, contour, and musical expertise. *Psychological Research*, 86(2), 421–442. <https://doi.org/10.1007/s00426-021-01519-0>
- Talamini, F., Carretti, B., & Grassi, M. (2016). THE WORKING MEMORY OF MUSICIANS AND NON MUSICIANS. 183–191.
- Talavage, T. M., Sereno, M. I., Melcher, J. R., Ledden, P. J., Rosen, B. R., & Dale, A. M. (2004). Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *Journal of neurophysiology*, 91(3), 1282-1296.
- Tamana, S. K., Victor, E., Joyce, C., Diana, L. L., Meghan B., A., Theo J., M., Padmaja, S., Allan B., B., Stuart E., T., Malcolm R., S., Bruce D., D., Valerie, C., Carmen, R., Investigators, C. study, Jacqueline, P., & Piush J., M. (2019). Association between Screen Time and Children’s Performance on a Developmental Screening Test. *PLoS ONE*, 14(4). <https://doi.org/10.1001/jamapediatrics.2018.5056>
- Tanna, R. J., Lin, J. W., & De Jesus, O. (2021). Sensorineural hearing loss. In *StatPearls* [Internet]. StatPearls Publishing.
- Taylor-Phillips, S., & Stinton, C. (2019). Fatigue in radiology: A fertile area for future research. *British Journal of Radiology*, 92(1099). <https://doi.org/10.1259/bjr.20190043>
- Teki, S., Barascud, N., Picard, S., Payne, C., Griffiths, T. D., & Chait, M. (2016). Neural correlates of auditory figure-ground segregation based on temporal coherence. *Cerebral Cortex*, 26(9), 3669–3680. <https://doi.org/10.1093/cercor/bhw173>
- Teki, S., Chait, M., Kumar, S., Shamma, S., & Griffiths, T. D. (2013). Segregation of complex acoustic scenes based on temporal coherence. *ELife*, 2013(2), 1–16. <https://doi.org/10.7554/eLife.00699>
- Teki, S., Chait, M., Kumar, S., Von Kriegstein, K., & Griffiths, T. D. (2011). Brain bases for auditory stimulus-driven figure-ground segregation. *Journal of Neuroscience*, 31(1), 164–171. <https://doi.org/10.1523/JNEUROSCI.3788-10.2011>

- Thiel, C. M., & Fink, G. R. (2007). Visual and auditory alertness: Modality-specific and supramodal neural mechanisms and their modulation by nicotine. *Journal of Neurophysiology*, 97(4), 2758–2768. <https://doi.org/10.1152/jn.00017.2007>
- Thompson, S. K., Carlyon, R. P., & Cusack, R. (2011). An Objective Measurement of the Build-Up of Auditory Streaming and of Its Modulation by Attention. *Journal of Experimental Psychology: Human Perception and Performance*, 37(4), 1253–1262. <https://doi.org/10.1037/a0021925>
- Tibo, M. (2020). Decoding EEG responses during perception and imagination of music.
- Tiego, J., Testa, R., Bellgrove, M. A., Pantelis, C., & Whittle, S. (2018). A hierarchical model of inhibitory control. *Frontiers in Psychology*, 9(AUG), 1–25. <https://doi.org/10.3389/fpsyg.2018.01339>
- Tierney, A., Rosen, S., & Dick, F. (2019). Speech-in-Speech Perception, Nonverbal Selective Attention, and Musical Training. *Journal of Experimental Psychology: Learning Memory and Cognition*. <https://doi.org/10.1037/xlm0000767>
- Torppa, R., & Huotilainen, M. (2019). Why and how music can be used to rehabilitate and develop speech and language skills in hearing-impaired children. *Hearing Research*, 380, 108–122. <https://doi.org/10.1016/j.heares.2019.06.003>
- Treisman, A. M. (1964). Selective attention in man. *British Medical Bulletin*, 20(1), 12–16.
- Treisman, A. M., & Gelade, G. (1980). A Feature-Integration Theory of Attention. *Cognitive Psychology*, 12, 97–136. <https://doi.org/10.1097/00075198-199608000-00002>
- Uhlir, M., Fairhurst, M. T., & Keller, P. E. (2013). The importance of integration and top-down salience when listening to complex multi-part musical stimuli. *NeuroImage*, 77, 52–61. <https://doi.org/10.1016/j.neuroimage.2013.03.051>
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37(3), 498–505.
- Unsworth, N., & McMillan, B. D. (2013). Mind wandering and reading comprehension: Examining the roles of working memory capacity, interest, motivation, and topic experience. *Journal of Experimental Psychology: Learning Memory and Cognition*, 39(3), 832–842. <https://doi.org/10.1037/a0029669>
- Unsworth, N., Redick, T. S., Lakey, C. E., & Young, D. L. (2010). Lapses in sustained attention and their relation to executive control and fluid abilities: An individual differences investigation. *Intelligence*, 38(1), 111–122. <https://doi.org/10.1016/j.intell.2009.08.002>
- Unsworth, N., & Robison, M. K. (2016). Pupillary correlates of lapses of sustained attention. *Cognitive, Affective and Behavioral Neuroscience*, 16(4), 601–615. <https://doi.org/10.3758/s13415-016-0417-4>
- Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the antisaccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning Memory and Cognition*, 30(6), 1302–1321. <https://doi.org/10.1037/0278-7393.30.6.1302>
- van Besouw, R. M., Nicholls, D. R., Oliver, B. R., Hodkinson, S. M., & Grasmeder, M. L. (2014). Aural rehabilitation through music workshops for cochlear implant users. *Journal of the American Academy of Audiology*, 25(04), 311–323.
- Van Lancker, D., & Kreiman, J. (1987). Voice discrimination and recognition are separate abilities. *Neuropsychologia*, 25(5), 829–834. [https://doi.org/10.1016/0028-3932\(87\)90120-5](https://doi.org/10.1016/0028-3932(87)90120-5)
- van Moorselaar, D., & Slagter, H. A. (2020). Inhibition in selective attention. *Annals of the New York Academy of Sciences*, 1464(1), 204–221. <https://doi.org/10.1111/nyas.14304>
- Van Noorden, L. P. A. S. (1975). Temporal coherence and the perception of temporal position in tone sequences. *IPO Annual Progress Report*, 10, 4–18.

- 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 Zuijen, T. L., Sussman, E., Winkler, I., Näätänen, R., & Tervaniemi, M. (2004). Grouping of Sequential Sounds - An Event-Related Potential Study Comparing Musicians and Non musicians. *Journal of Cognitive Neuroscience*, 16(2), 331–338. <https://doi.org/10.1162/089892904322984607>
- Vandecapelle, S., Deckers, L., Das, N., Ansari, A. H., Bertrand, A., & Francart, T. (2020). EEG-based detection of the attended speaker and the locus of auditory attention with convolutional neural networks. *BioRxiv*, 475673. <https://doi.org/10.1101/475673>
- Vanden Bosch der Nederlanden, C. M., Joannisse, M. F., & Grahn, J. A. (2020). Music as a scaffold for listening to speech: Better neural phase-locking to song than speech. *NeuroImage*, 214(May 2019), 116767. <https://doi.org/10.1016/j.neuroimage.2020.116767>
- Vanden Bosch der Nederlanden, C. M., Joannisse, M. F., Grahn, J. A., Snijders, T. M., & Schoffelen, J. M. (2022). Familiarity modulates neural tracking of sung and spoken utterances. *NeuroImage*, 252(August 2021), 119049. <https://doi.org/10.1016/j.neuroimage.2022.119049>
- Vandierendonck, A. (2021). On the utility of integrated speed-accuracy measures when speed-accuracy trade-off is present. *Journal of Cognition*, 4(1), 1–26. <https://doi.org/10.5334/joc.154>
- Verbruggen, F., & Logan, G. D. (2009). Models of response inhibition in the stop-signal and stop-change paradigms. *Neuroscience and Biobehavioral Reviews*, 33(5), 647–661. <https://doi.org/10.1016/j.neubiorev.2008.08.014>
- Verhaeghen, P., & Basak, C. (2005). Aging and switching of the focus of attention in working memory: Results from a modified N-Back task. *The Quarterly Journal of Experimental Psychology Section A*, 58(1), 134–154.
- Verschueren, E., Vanthornhout, J., & Francart, T. (2020). The effect of stimulus choice on an EEG-Based objective measure of speech intelligibility. In *Ear and Hearing* (Issue 637424). <https://doi.org/10.1097/AUD.0000000000000875>
- Vidal, J. J. (1973). Toward direct brain-computer communication. *Annual Review of Biophysics and Bioengineering*, 2, 157–180. <https://doi.org/10.1146/annurev.bb.02.060173.001105>
- Vince, M. A. (1948). The intermittency of control movements and the psychological refractory period. *British Journal of Psychology*, 38(3), 149.
- Vliegen, J., Moore, B. C. J., & Oxenham, A. J. (1999). The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. *The Journal of the Acoustical Society of America*, 106(2), 938–945. <https://doi.org/10.1121/1.427140>
- Vliegen, J., & Oxenham, A. (1999). Sequential stream segregation in the absence of spectral cues. *Journal of the Acoustical Society of America*, 105(1), 339–346.
- von Békésy, G., & Peake, W. T. (1990). Experiments in Hearing. *The Journal of the Acoustical Society of America*, 88(6), 2905–2905. <https://doi.org/10.1121/1.399656>
- von Helmholtz, H. L. F. (2004). Some laws concerning the distribution of electric currents in volume conductors with applications to experiments on animal electricity. *Proceedings of the IEEE*, 92(5), 868–870.
- Wang, H., & Fan, J. (2007). Human attentional networks: A connectionist model. *Journal of Cognitive Neuroscience*, 19(10), 1678–1689. <https://doi.org/10.1162/jocn.2007.19.10.1678>
- Wang, Y., Zhang, J., Zou, J., Luo, H., & Ding, N. (2019). Prior Knowledge Guides Speech Segregation in Human Auditory Cortex. *Cerebral Cortex*, 29(4), 1561–1571. <https://doi.org/10.1093/cercor/bhy052>
- Warr, W. B. (1992). Organization of olivocochlear efferent systems in mammals. *The mammalian auditory pathway: Neuroanatomy*, 410–448.

- Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology Section A*, 49(1), 51-79.
- Watzl, S. (2017). *Structuring mind: The nature of attention and how it shapes consciousness*. Oxford University Press.
- Weineck, K., Wen, O. X., & Henry, M. J. (2022). Neural entrainment is strongest to the spectral flux of slow music and depends on familiarity and beat salience. *ELife*. <https://www.biorxiv.org/content/10.1101/2021.11.29.470396v2%0Ahttps://www.biorxiv.org/content/10.1101/2021.11.29.470396v2.abstract>
- Wiecki, T. V., & Frank, M. J. (2013). A computational model of inhibitory control in frontal cortex and basal ganglia. *Psychological Review*, 120(2), 329–355. <https://doi.org/10.1037/a0031542>
- 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. *Journal of Neuroscience*, 32(40), 14010–14021. <https://doi.org/10.1523/JNEUROSCI.1528-12.2012>
- Williams, D. A. H., & Miranda, E. R. (2018). BCI for Music Making: Then, Now, and Next. In *Brain-Computer Interfaces Handbook* (pp. 193–206). CRC Press.
- Williges, B., Wesarg, T., Jung, L., Geven, L. I., Radeloff, A., & Jürgens, T. (2019). Spatial Speech-in-Noise Performance in Bimodal and Single-Sided Deaf Cochlear Implant Users. *Trends in Hearing*, 23, 1–16. <https://doi.org/10.1177/2331216519858311>
- Wilson, B. S., & Dorman, M. F. (2018). A Brief History of the Cochlear Implant and Related Treatments. In *Neuromodulation (Second Edi)*. Elsevier Ltd. <https://doi.org/10.1016/b978-0-12-805353-9.00099-1>
- Winkler, I., Takegata, R., & Sussman, E. (2005). Event-related brain potentials reveal multiple stages in the perceptual organization of sound. *Cognitive Brain Research*, 25(1), 291–299. <https://doi.org/10.1016/j.cogbrainres.2005.06.005>
- Wong-Lin, K., Philip Eckhoff, Holmes, P., & Cohen, J. D. (2010). Optimal performance in a countermanding saccade task. *Brain Research*, 178–187. <https://doi.org/10.1016/j.brainres.2009.12.018>. Optimal
- Wong, D. D. E., Fuglsang, S. A., Hjortkjær, J., Ceolini, E., Slaney, M., & de Cheveigné, A. (2018). A comparison of regularization methods in forward and backward models for auditory attention decoding. *Frontiers in Neuroscience*, 12(AUG), 1–16. <https://doi.org/10.3389/fnins.2018.00531>
- Woods, D. L., Herron, T. J., Cate, A. D., Yund, E. W., Stecker, G. C., Rinne, T., & Kang, X. (2010). Functional properties of human auditory cortical fields. *Frontiers in Systems Neuroscience*, 4(December), 1–13. <https://doi.org/10.3389/fnsys.2010.00155>
- Woods, D. L., Stecker, G. C., Rinne, T., Herron, T. J., Cate, A. D., Yund, E. W., Liao, I., & Kang, X. (2009). Functional maps of human auditory cortex: Effects of acoustic features and attention. *PLoS ONE*, 4(4). <https://doi.org/10.1371/journal.pone.0005183>
- Wright, J., & Bregman, A. S. (1987). Auditory stream segregation and the control of dissonance in polyphonic music. In *Contemporary Music Review* (Vol. 2, Issue 1). <https://doi.org/10.1080/07494468708567054>
- Wu, D., Li, C., Yin, Y., Zhou, C., & Yao, D. (2010). Music composition from the brain signal: Representing the mental state by music. *Computational Intelligence and Neuroscience*, 2010(March). <https://doi.org/10.1155/2010/267671>
- Wu, W. (2014). *Attention*. Routledge.
- Wühr, P., Frings, C., & Heuer, H. (2018). Response preparation with reliable cues decreases response competition in the Flanker task. *Experimental psychology*, 65(5), 286.
- Wurm, L. H., Labouvie-Vief, G., Aycock, J., Rebucal, K. A., & Koch, H. E. (2004). Performance in auditory and visual emotional stroop tasks: A comparison of older and younger adults. *Psychology and Aging*, 19(3), 523–535. <https://doi.org/10.1037/0882-7974.19.3.523>

- Yanko, M. R., & Spalek, T. M. (2014). Driving with the wandering mind: The effect that mind-wandering has on driving performance. *Human Factors*, 56(2), 260–269. <https://doi.org/10.1177/0018720813495280>
- Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 10(5), 601–621. <https://doi.org/10.1037/0096-1523.10.5.601>
- Yin, T. C., & Chan, J. C. (1990). Interaural time sensitivity in medial superior olive of cat. *Journal of neurophysiology*, 64(2), 465–488.
- Yonan, C. A., & Sommers, M. S. (2000). The effects of talker familiarity on spoken word identification in younger and older listeners. *Psychology and aging*, 15(1), 88.
- Yoo, J., & Bidelman, G. M. (2019). Linguistic, perceptual, and cognitive factors underlying musicians' benefits in noise-degraded speech perception. *Hearing Research*, 377, 189–195. <https://doi.org/10.1016/j.heares.2019.03.021>
- Zahnert, T. (2011). The Differential Diagnosis of Hearing Loss. *Deutsches Ärzteblatt International*, 108(25), 433–444. <https://doi.org/10.3238/arztebl.2011.0433>
- Zander, T. O., & Kothe, C. (2011). Towards passive brain-computer interfaces: Applying brain-computer interface technology to human-machine systems in general. *Journal of Neural Engineering*, 8(2). <https://doi.org/10.1088/1741-2560/8/2/025005>
- Zarate, J. M., Ritson, C. R., & Poeppel, D. (2012). Pitch-interval discrimination and musical expertise: Is the semitone a perceptual boundary? *The Journal of the Acoustical Society of America*, 132(2), 984–993. <https://doi.org/10.1121/1.4733535>
- Zatorre, R. J. (1988). Pitch perception of complex tones and human temporal-lobe function. *Journal of the Acoustical Society of America*, 84(2), 566–572.
- Zatorre, R. J. (1979). Recognition of dichotic melodies by musicians and non musicians. *Neuropsychologia*, 17(6), 607–617. [https://doi.org/10.1016/0028-3932\(79\)90035-6](https://doi.org/10.1016/0028-3932(79)90035-6)
- Zatorre, R. J., & Belin, P. (2001). Spectral and Temporal Processing in Human Auditory Cortex. *Cerebral Cortex*, 11(10), 946–953. <https://doi.org/10.1093/cercor/12.2.140>
- Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, 6(1), 37–46. [https://doi.org/10.1016/S1364-6613\(00\)01816-7](https://doi.org/10.1016/S1364-6613(00)01816-7)
- Zekveld, A. A., Kramer, S. E., van Steenbergen, H., van Scheepen, J. A. M., & Versfeld, N. J. (2020). The influence of hearing loss on cognitive control in an auditory conflict task: Behavioral and pupillometry findings. *Journal of Speech, Language, and Hearing Research*, 63(7), 2483–2492. https://doi.org/10.1044/2020_JSLHR-20-00107
- Zelazo, P. D., Blair, C., & Willoughby, M. T. (2017). Executive Function: Implications for Education. National Center for Education Research. <http://ies.ed.gov/>.
- Zendel, B. R., & Alain, C. (2013). The influence of lifelong musicianship on neurophysiological measures of concurrent sound segregation. *Journal of Cognitive Neuroscience*, 25(4), 503–516. https://doi.org/10.1162/jocn_a_00329
- Zendel, B. R., & Alain, C. (2009). Concurrent sound segregation is enhanced in musicians. *Journal of Cognitive Neuroscience*, 21(8), 1488–1498. <https://doi.org/10.1162/jocn.2009.21140>
- Zendel, B. R., Tremblay, C.-D., Belleville, S., & Peretz, I. (2015). The Impact of Musicianship on the Cortical Mechanisms Related to Separating Speech from Background Noise. *Journal of Cognitive Neuroscience*, 27(15), 1044–1059. <https://doi.org/10.1162/jocn>
- Zeng, F. G., Rebscher, S., Harrison, W., Sun, X., & Feng, H. (2008). Cochlear Implants: System Design, Integration, and Evaluation. *IEEE Reviews in Biomedical Engineering*, 1(dc), 115–142. <https://doi.org/10.1109/RBME.2008.2008250>

Zhang, H., & Kornblum, S. (1999). A Parallel Distributed Processing Model of Stimulus – Stimulus and Stimulus – Response Compatibility. *Cognitive Psychology*, 38, 386–432.

Zink, R., Baptist, A., Bertrand, A., Van Huffel, S., & De Vos, M. (2016). Online detection of auditory attention in a neurofeedback application. *Proc. 8th International Workshop on Biosignal Interpretation*, 1–4.

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., Goodman, R. R., Emerson, R., Mehta, A. D., Simon, J. Z., Poeppel, D., & Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party.” *Neuron*, 77(5), 980–991. <https://doi.org/10.1016/j.neuron.2012.12.037>

Zoefel, B., & VanRullen, R. (2015). Selective perceptual phase entrainment to speech rhythm in the absence of spectral energy fluctuations. *Journal of Neuroscience*, 35(5), 1954–1964. <https://doi.org/10.1523/JNEUROSCI.3484-14.2015>

Zuk, J., Benjamin, C., Kenyon, A., & Gaab, N. (2014). Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PLoS ONE*, 9(6). <https://doi.org/10.1371/journal.pone.0099868>