

UiO : **University of Oslo**

Merve Akça

Attending to Sounds in the Blink of an Eye

Thesis submitted for the degree of Philosophiae Doctor

Department of Musicology
Faculty of Humanities

RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion



2022

© Merve Akça, 2022

*Series of dissertations submitted to the
Faculty of Humanities, University of Oslo*

All rights reserved. No part of this publication may be reproduced or transmitted, in any form or by any means, without permission.

Cover: Hanne Baadsgaard Utigard.
Print production: 07-Media Oslo.

*To my nieces Sira, Mina, and Filippa -
may you never stop asking questions.*

Abstract

When we attend to something within a rapid sequence, the ‘eyes of attention’ appear to shut for a short period of time, limiting our information processing ability. This thesis details an investigation into the factors that could potentially modulate the limits of perception and attention to brief sounds over time. Due to a scarcity of research and unresolved debates, the temporal limits of perception and attentional selection in the auditory modality are not sufficiently understood. In a series of empirical studies, we tested whether these temporal limits can be modulated through various experimental manipulations and/or individual differences. These studies yielded a complex pattern of results which suggests that human voices can effectively overcome these limits. This is in line with the evolutionary and neuroscientific perspectives highlighting the importance of human voice perception for humans. However, this modulatory effect of sound category appears to vary depending on the duration that the sounds are presented for. The investigations of factors related to individual differences (i.e., musicality, working memory, and impulsivity) were not conclusive, at least within the paradigms employed here. In addition to the behavioural findings, pupillometry data shed light into the amount of mental effort linked with attending to brief sounds in time. The current work has contributed to, and is in line with, the recent direction of the empirical findings and theoretical accounts of temporal selective attention. It has also offered several directions for future research with important theoretical and societal implications.

Sammendrag

Når vi identifiserer noe i en rask sekvens, virker det som at 'oppmerksomhetens øyne' lukkes i en kort periode, hvilket begrenser vår evne å prosessere informasjon. Denne avhandling undersøker faktorer som potensielt kan modulere begrensninger i persepsjon og oppmerksomhet til kortvarige lyder over tid. På grunn av en begrenset mengde forskning og uavklarte debatter innenfor forskningsfeltet, har de tidsmessige begrensningene i auditiv persepsjon og oppmerksomhet ikke nådd tilstrekkelig klarhet. I en serie av empiriske studier, testet vi hvorvidt disse begrensningene kan moduleres gjennom diverse eksperimentelle manipulasjoner og individuelle forskjeller. Disse studiene ga et komplekst mønster av resultater hvilke indikerte at menneskestemmer ikke synes å påvirkes av disse begrensningene. Dette er i tråd med evolusjonære og nevrovitenskaplige perspektiver som belyser viktigheten av persepsjon av menneskestemmer for mennesker. Denne modulerende effekten av lyd kategorier synes imidlertid å variere basert på varigheten av lydene som presenteres. Undersøkelsene av faktorer knyttet til individuelle forskjeller (for eksempel, musikalitet, arbeidshukommelse og impulsivitet) ga ikke konkluderende resultater, i hvert fall ikke med paradigmen som ble benyttet her. I tillegg til de adferdsmessige funnene, kunne data fra pupillometri belyse hvor mye mental anstrengelse som trengs for å oppfatte kortvarige lyder. Denne avhandling har bidratt til den nylige retningen av empiriske funn og teorier om tidsmessig selektiv oppmerksomhet. Den foreslår også flere retninger for fremtidig forskning med viktige teoretiske og samfunnsmessige implikasjoner.

Acknowledgements

This work has challenged me in many ways, yet at the same time given me unique opportunities to explore, experiment, make mistakes, and dare to show up. I am grateful for every moment of this PhD, but most importantly, for the people who supported me and believed in me throughout this journey.

First and foremost, I would like to express my gratitude for my supervisors and co-authors Jonna Vuoskoski, Rolf Inge Godøy, Laura Bishop, and Bruno Laeng for sharing their invaluable knowledge, and for their feedback and guidance in the pursuit of this PhD project. Many thanks to all four!

A big thanks to Anne Danielsen for her compassionate leadership at RITMO that permits people to grow and for her continuous support throughout my PhD journey. I extend my gratitude to Alexander Jensenius for co-creating an inspiring research environment and for his encouraging words that led me to apply to this doctoral position in the first place! Many thanks also to Zafer Özgen and Peter Edwards for their excellent leadership at IMV. Thanks so much Anne-Cathrine Wesnes and also Målfrid Hoaas, Marit Furunes, and Eirik Karlsen for all their kindness and administrative help. Thanks to the Research Council of Norway for their financial support of RITMO's projects and to University of Oslo for providing excellent conditions for scientific research.

Thanks Tor Endestad for letting me pick your brain about some of the methodological concerns. Thanks for much appreciated assistance during data collection Vegard Volehaugen, Emilie Holte and Sumeyya Atmaca. I would also like to thank all the participants who volunteered for the studies in this thesis. Your participation made this work possible!

Being surrounded by friendly, smart, and inspiring colleagues made my everyday life a lot more pleasant. Thanks to all past and present colleagues at RITMO and IMV for the stimulating discussions! To name but a few of my past and present fellow PhD and Post-doc researchers Agata, Marek, Victor, Gui, Mari, Maja, Emil, Connor, Henrik, Tejaswinee, Çağrı, Julian, Dana, Alex. It was a pleasure to get to know you and talk to you just about anything.

I am forever grateful to my therapist Burçin and my mindfulness community for creating a safe space in the stormy days and for sitting with me. My mother-in-law Titti, tack för att jag fått låna ditt fina kontor, where I really enjoyed writing article drafts with the furry and cosy company of Molly o Sarek. My dear parents, canım annem ve babam, üzerimdeki emekleriniz için size ne kadar teşekkür etsem az. You taught me that hard work pays off, always. Last, but absolutely not least, Marcus, min älskling, din kärlek och ditt ovillkorliga stöd uppskattas mycket. I am forever grateful for having you by my side.

List of papers

Paper I

Akça, M. Laeng, B. and Godøy, R. I. 'No evidence for an auditory attentional blink for voices regardless of musical expertise'. In: *Frontiers in Psychology* **10** (2020), 2935. DOI: 10.3389/fpsyg.2019.02935.

Paper II

Akça, M. Vuoskoski, J. K. Laeng, B. and Bishop, L. 'Recognition of brief sounds in rapid serial auditory presentation'. *Submitted for publication*.

Paper III

Akça, M. Bishop, L. Vuoskoski, J. K. and Laeng, B. 'What does the pupil whisper? Tracing the temporal costs of auditory selective attention with pupillometry'. Under review in *Frontiers in Neuroscience*.

Datasets can be accessed here: <https://osf.io/ws2xk/>.

Abbreviations

AB	Attentional blink
ANOVA	Analysis of variance
BF	Bayes factor
EEG	Electroencephalography
ERP	Event-related potential
fMRI	Functional magnetic resonance imaging
ISI	Inter stimulus interval
LC-NA	Locus coeruleus noradrenaline
MEG	Magnetoencephalography
ms	Millisecond
rm	Repeated measures
RMS	Root mean square
RSAP	Rapid serial auditory presentation
RSVP	Rapid serial visual presentation
RT	Response time
SD	Standard deviation
SE	Standard error of the mean
SMI	SensoMotoric Instruments
SOA	Stimulus onset asynchrony
STS	Superior temporal sulcus
T1	First target stimulus
T2	Second target stimulus
T2 T1	Accuracy of the second target conditional on the correct report of the first target
WM	Working memory

Contents

Abstract	iii
Sammendrag	v
List of papers	ix
Abbreviations	xi
Contents	xiii
1 Introduction	1
1.1 Objectives	2
1.2 Research questions	2
1.3 Operational definitions	3
1.4 Scope of the thesis	4
1.5 Outline of the thesis	5
2 Voice cognition	7
2.1 Introduction	7
2.2 Evolutionary perspectives on voice cognition	8
2.2.1 Adaptive problems related to voice	9
2.2.2 Ontogenetic origins of voice cognition	10
2.2.3 Voice cognition among nonhuman species	10
2.3 Behavioural and neuroscientific perspectives on human voice cognition	12
2.3.1 Human voice areas in the brain	12
2.3.2 Are human voices processed more quickly than other auditory stimuli?	13
2.3.3 Are human voices detected, categorized, and recognized better than other auditory stimuli?	14
2.3.4 Similarities and differences between voice and face cognition	15
3 Temporal limits of attention	19
3.1 Introduction	19
3.2 Dual-task costs in rapid serial presentation paradigm	20
3.3 Attentional blink phenomenon	21
3.3.1 The auditory attentional blink	22
3.3.2 Theories of attentional blink	25
3.3.3 Locus coeruleus-noradrenaline system	26
	xiii

3.4	Modulating the attentional blink	28
3.4.1	Tasks we ‘blink’ less to	29
3.4.2	Stimuli we ‘blink’ less to	29
3.4.3	Individuals who do not ‘blink’ or ‘blink’ less . . .	31
3.4.4	Modulating the auditory attentional blink	33
3.5	Musicality and musical expertise	34
3.5.1	Benefits on music-related abilities	35
3.5.2	Benefits for cognitive functions	36
3.5.3	Musicians and the limits of attention	37
4	Methodology	39
4.1	Experimental approaches	39
4.1.1	Rapid serial auditory presentation paradigm . . .	39
4.1.2	Auditory attentional blink task	40
4.2	Behavioural measures	41
4.2.1	Target accuracy	41
4.2.2	Target recognition sensitivity	41
4.2.3	Maximal AB and AB magnitude	42
4.2.4	Response time	42
4.3	Pupillary measures	43
4.4	Working memory measures	46
4.5	Self-report measures	47
4.5.1	Musical sophistication	47
4.5.2	Impulsivity	47
4.5.3	Subjective certainty ratings	48
4.6	Participant recruitment and sample characteristics	48
4.7	Experimental design considerations	49
4.8	Statistical approach and data analysis	52
4.8.1	Methods of statistical inference	52
4.8.2	Bayesian approach to statistics	53
4.8.3	Statistical considerations	55
4.8.4	Data analysis	57
5	Research summary	59
5.1	Paper I	59
5.1.1	Abstract	59
5.1.2	Key findings	60
5.1.3	Open questions	62
5.2	Paper II	62
5.2.1	Abstract	62
5.2.2	Key findings	63
5.2.3	Open questions	64
5.3	Paper III	65
5.3.1	Abstract	65
5.3.2	Key findings	66
5.3.3	Open questions	67

6	General discussion	69
6.1	Introduction	69
6.2	Summary and discussion of research contributions	70
6.3	Limitations and Strengths	74
6.3.1	Limitations	74
6.3.2	Strengths	75
6.4	Future directions	76
	Bibliography	79
	Papers	98
I	No evidence for an auditory attentional blink for voices regardless of musical expertise	99
II	Recognition of brief sounds in rapid serial auditory presentation	115
III	What does the pupil whisper? Tracing the temporal costs of auditory selective attention with pupillometry	147
	Appendices	177
A	Bayesian correlational tests for Papers II and III	179
B	Bayes Factor evidence categories	181

Chapter 1

Introduction

Within all the cacophony of sounds we hear in everyday life, what are our limits in processing auditory information? Consider a situation in which you are to listen to a multitude of sounds, but among these you only want to attend to certain sound events. How well do you think you can catch these ‘target sounds’? What if they are very brief and presented in close temporal proximity? Would you still be able to catch all targets?

At the core of this dissertation lies the question: What are some of the conditions that can potentially modulate these limits in how we distribute our attention to the sounds over time? In further pursuit of unravelling these rather important, yet not fully answered questions, this dissertation addresses the overarching topic of temporal limits of human auditory cognition. In particular, it focuses on the temporal limits of human perception and attentional selection processes for brief human voices presented among rapid sound sequences.

As the events in our environment unfold in time, our attentional systems do not always enable us to attend to them, but rather vary in order to optimize the selection and prioritization of events happening at a particular time. These temporal constraints on the human attentional system could be seen both as a curse and a blessing. On the one hand, if we could attend to everything happening at all times, we would be too overwhelmed by irrelevant information to act upon our goals. On the other hand, if attention is inherently constrained no matter what happens during a specific time window, we could miss critical information.

An example of such an occurrence is the attentional blink (AB) phenomenon, where the ‘eyes of attention’ seem to be shut for a brief duration. More specifically, this is a robust observation of an attentional cost lasting for about half a second. As we will explore in the rest of this thesis, this puzzling phenomenon suggests that, under certain conditions, people tend to temporarily fail to attend to information presented to them. This robust effect (i.e., a failure in selection within this critical time window) is often considered to reflect the limits of the human attentional system. Note, however, that much of what we know about this phenomenon is heavily based on the visual domain. Thus, to approach a fuller picture and to have a better understanding of the so-called limits of human cognition, it is important to study this complex phenomenon in the auditory modality, as well as in other sensory modalities.

A starting point in this dissertation is the question of whether these temporal limitations of the human attentional system are unavoidable. If the constraints on our attentional and perceptual system function to optimize the cognitive resources, it would be neither optimal nor good for our survival if there would be no way to ‘bend the rules’ in critical situations and/or in response to critical

1. Introduction

events. It is only recently that researchers interested in the temporal dynamics of attention in the visual modality started to explore the situations where these limits no longer apply. In this work, I seek to understand what these situations might be in the auditory modality. To further our understanding of the temporal limits of human cognition in the auditory modality, I will refer to a series of empirical studies (i.e., Papers I, II, and III) that were conducted in the RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion at University of Oslo during my doctoral fellowship.

1.1 Objectives

The primary objective of this thesis is to investigate potential modulations of the temporal limits of auditory selective attention and perception through stimulus-related factors. All investigations presented in this thesis aimed to examine whether human voices preferentially recruit the limited-capacity cognitive processes. Paper I aimed to investigate the temporal limits in the deployment of attention for auditory objects related to common perceptual and musical expertise. Paper II aimed at determining the effects of timbre and duration on the recognition of brief sounds embedded in a rapid sound sequence. Paper III aimed to study the mental effort associated with attentional selection of human voices (as compared with other sounds) under two stimulus duration conditions.

A secondary objective of the present thesis is to shed light on the complex relationship between the temporal limits of human auditory cognition and participant-related factors (e.g., musicality, expertise, working memory, etc.). In line with this objective, Paper I set out to investigate the potential benefits of musical expertise and common perceptual expertise on temporal selective auditory attention. It also explored, at an individual level, the relationship between the size of attentional limitations, working memory, and musicality. In Paper II we explored how individual differences in participants' musicality and working memory scores may relate to the recognition of brief sounds (as well as various other dependent measures, e.g., response time). Finally, Paper III investigated the link between individual differences in musicality and trait impulsivity, and the size of the limitations in selective attending to auditory targets.

1.2 Research questions

There are three main research questions in this thesis. The first research question is: can stimulus-related factors modulate the temporal limits of auditory selective attention and perception? Four sub-questions have emerged from this question:

1. Do human voices (and other sounds that share similarity to human voices) preferentially recruit temporal selective attention processes, and hence escape (or suffer less from) the temporal limits of auditory attention? In other words, are human voices 'capacity-free' or do they require less attentional resources compared with other sounds?

2. Are brief human voices recognized better compared with other sounds when presented among environmental sounds?
3. What is the role of sound duration in recognizing and in selectively attending to very brief (at millisecond level) sounds presented amongst a rapid sound sequence?
4. Does duration influence the recognition of and attention to human voices and other timbres similarly or differently? In other words, is less information sufficient for voices to be recognized and attentionally selected in complex listening situations?

The second research question is: what are the some of the participant-related factors that can modulate the temporal limits of auditory attention and perception? It has three sub-questions:

1. Do musicality or expertise offer any advantage in temporal selective auditory attention and in recognition of brief sounds?
2. Is working memory related to the magnitude of the failures in temporal selective auditory attention and in recognition of brief sounds?
3. Are personality traits that are linked to information processing, such as impulsivity, related to the temporal selective auditory attention?

The third and final research question is: whether psychophysiological measure of mental effort (i.e., pupil dilation) can throw light on the temporal limits of auditory selective attention? The main sub-question that emerged from this methodological question was:

1. If we can modulate the temporal limits of auditory selective attention through our experimental manipulations, would the fluctuations in pupillary traces reflect the mental effort associated with these limits of auditory selective attention?

1.3 Operational definitions

Despite the regular use of the words ‘attention’, ‘expertise’ and ‘musicality’ in everyday discourse, their formal definitions are rather elusive. Scientifically, all three are known to be not unitary, but rather complex and multifaceted constructs. Despite many scholarly attempts towards demystifying the confusion to improve the scientific communication, all three, in fact also ‘working memory’, are difficult to operationalise under one common definition. This is understandable given the variations in the applications within different disciplines and the complexity of each of these constructs.

In the course of this thesis, I will be referring to a particular aspect of attention called the *temporal selective attention*. The selective aspect refers to the process through which some informational units are prioritised over others,

1. Introduction

and the temporal aspect means that this process operates in a time-based frame of reference (Cohen, 2014). As for the expertise and musicality, we have used (see Paper I) the definitions of: “advanced levels of music performance experience combined with extensive training in music” to refer to *musical expertise*, and “common human experience associated with higher level of perceptual capacity for voices, shaped by extensive exposure to human voices for years on a daily basis” to refer to *perceptual expertise with voices*. These definitions were formulated in relation to the research questions in Paper I and affected how we recruited participants (I will be explaining this further in Chapter 4). *Musicality* was treated more as a spectrum among general population. Since musicality was measured through a musical sophistication index in this thesis, I will operationally define it in the same way as *musical sophistication*, which refers to “a psychometric construct that can refer to musical skills, expertise, achievements, and related behaviours across a range of facets” (Müllensiefen et al., 2014b, p. 2).

Finally, I use *working memory* (WM) in the same way as in the multi-component WM system (Baddeley & Hitch, 1974), and refer to WM not as a passive storage, but as a combination of storage and manipulation. Auditory WM should then be understood not just a process of passively maintaining sounds in our mind over short periods of time, but also simultaneously processing and manipulating them in our ongoing mental activities. Another aspect of WM that is worth mentioning here is whether auditory WM for verbal and tonal material should be understood as two independent WM systems. Unfortunately, the studies directly comparing auditory WM for different material is limited. Schulze, Koelsch & Williamson (2018) summarizes the results from the behavioural studies as inconsistent but the few comparison studies using neuroimaging seem to indicate a large overlap of neural networks underlying WM for verbal and tonal information.

1.4 Scope of the thesis

The scholarly contributions that attempt to explain the limitations of temporal selective attention, although mostly based on the findings from visual attention studies, are plentiful. In this work, I will not put forward new theories to explain these limitations, nor will I be able to fully explain the underlying cause of the findings in the present investigations. This would not be realistic, especially considering the fact that the investigations presented in this thesis are not designed to test any specific theory per se. Instead, my endeavour will be to bring together insights from past and present works.

This does not entail, however, that this thesis makes no theoretical contributions. Within the scope of this thesis, I will be exploring ways in which we can potentially modulate these limits through experimental manipulations and/or differences among individuals. As we will discover in Chapter 3, this approach is in line with the recent shifts in the theoretical understanding of the topic at hand. Theoretically, my starting point will be to find out what evidence we currently have, and what evidence will come out of the investigations in the

auditory modality in favour of (or against) these shifts. I will also interpret the findings in light of relevant theoretical arguments and empirical studies. As the findings from the auditory modality accumulate, more replication and meta-analysis studies will emerge and only then can we begin to approach a solid theoretical understanding of the issue at hand.

My work and the ideas behind this work mostly derive from, but are not limited to, cognitive psychology, cognitive neuroscience, systematic musicology, and neurobiology disciplines, and uses an evolutionary psychology approach to human auditory cognition. A selection of relevant literature from various disciplines will be covered to provide wider perspectives and a theoretical understanding relevant to the present work. As mentioned earlier, the concepts addressed in this thesis are multifaceted, and so is their literature. Therefore, it should be clear from the beginning that the reviews in the background chapters of this thesis (Chapters 2 and 3) are not complete (i.e., covering all the existing studies and theories in the literature). In the interest of brevity and structure, I have chosen some representative studies directly addressing the questions at hand. The selection was based on relevant papers I identified over the course of my PhD as a result of systematic search hits on various databases with the keywords addressing each question. I also consulted existing review papers on the topics at hand. These sources were then organized and revised either thematically or by argument to explain some of the wider issues behind this thesis.

In terms of research methodology, I see myself as a cognitive scientist who primarily uses methods from behavioural sciences and cognitive psychology. In the work presented in this thesis, I use behavioural and psycho-physiological (i.e., pupillometry) ways of exploring human auditory cognition. Complementary to what is presented here, if one wishes to explore the time-course of the attentional and perceptual processes measured from the participants' brain activity, electroencephalography (EEG) would be the most suitable method. Though I am not using this method, in the next chapter, I will be referring to previous studies using EEG, among other brain imaging methods, to bring also neuroscientific perspectives to the discussion.

1.5 Outline of the thesis

The thesis consists of the five chapters described below, followed by the three research papers that have been either published or submitted to scientific journals. Figure 1.1 illustrates the core concepts explored in the papers presented in this thesis.

Chapter 2: Voice cognition. This background chapter presents an overview of voice processing, which is one of the central concepts in this thesis.

Chapter 3: Temporal limits of attention. This chapter addresses relevant theories and studies of temporal limits of attention while positioning this topic in the auditory modality. Importantly, it presents the situations in which it is possible to overcome these limits. It also explores the research on the aspects

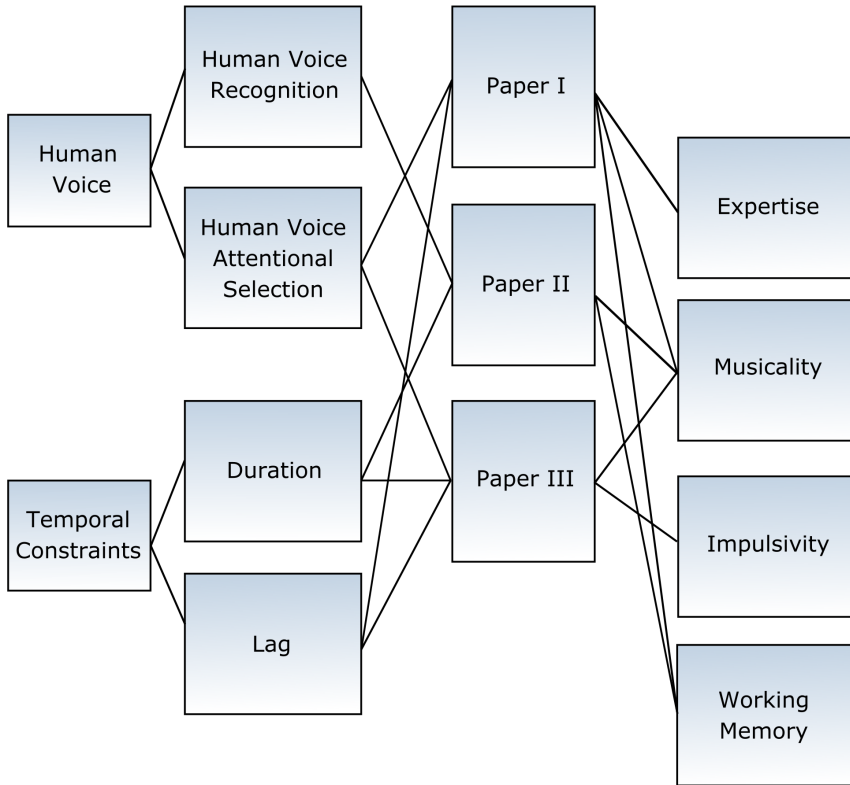


Figure 1.1: An illustration of the core concepts explored in the three research papers included in this thesis.

of expertise, musicality and working memory, with the goal of illustrating the complex relationship between temporal auditory attention and these aspects.

Chapter 4: Methodology. This chapter introduces the experimental approaches and the measures, while providing the rationale for employing them to investigate the auditory perceptual and attentional processes. It also explains and critically reflects on the details of the methodological and statistical aspects of this work, and summarises the processing and analysis of the data.

Chapter 5: Research summary. The summary chapter gives an overview of the findings of the investigations included in this thesis and puts forward some of the open questions raised by each investigation.

Chapter 6: General discussion. In this final chapter, I discuss and integrate the findings of the investigations included in this thesis, discuss this work in the light of the empirical studies and theoretical background introduced earlier, and propose directions for future research.

Chapter 2

Voice cognition

“We constantly inhabit the universe of voices, we are continuously bombarded by voices, we have to make our daily way through a jungle of voices, and we have to use all kinds of machetes and compasses so as not to get lost. There are the voices of other people, the voices of music, the voices of media, our own voice intermingled with the lot. [...] All those voices rise over the multitude of sounds and noises, another even wilder and wider jungle: sounds of nature, sounds of machines and technology. Civilization announces its progress by a lot of noise, and the more it progresses the noisier it gets.”

— Mladen Dolar

A common aim in all three papers in this thesis was to investigate whether human perception and selective auditory attention mechanisms are preferentially recruited by human voices. In this chapter, I will be positioning the significance of voices within a broader understanding of voice processing from evolutionary and neuroscientific perspectives¹.

2.1 Introduction

Voice is a ubiquitous and perhaps the most important sound in our daily environments. The voice can serve as many things: a language conveyor, a social bond, a highly flexible musical instrument, a body index, an identity register, a key component of acting, and a gauge of emotion (Smith, 2008). As it conveys significant information about a person, the voice has been likened to a fingerprint (Dolar, 2006), and to an auditory face (Belin et al., 2004).

First, a mere matter of terminology. In this thesis, voice refers to basic, audible sound produced by vocal fold vibration. I will use the term voice

¹These are not to be seen as alternatives (Nesse, 2019), but as an attempt toward a unification of evolutionary and proximate explanations underlying this issue.

2. Voice cognition

production to refer to the physical act of generating voice by humans and other vertebrates. This definition of voice distinguishes voice from speech, which allows for studying voice perception in paralinguistic sense. Speech, one of the primary manifestations of interpersonal communication among humans, is also supported through voice. In addition to, but also in the absence of speech (e.g., a baby’s cry, a young girl’s scream, a man clearing his throat), voice is rich in information that is socially relevant to humans (Belin, 2019). Belin (2019) defines the term voice cognition as “the set of auditory cognitive abilities, including speech perception, allowing us to extract information from vocal sounds as a particular sound category”. In this thesis, I use this term to refer to abilities to perceive, recognize and attend to voices among other sounds.

Voice production involves the interaction of two basic components: a source of acoustic energy (i.e., larynx) and a filter (i.e., oral and nasal cavities of the vocal tract located above the larynx) (Ghazanfar & Rendall, 2008). Figure 2.1 depicts the anatomy of these basic components and the surrounding organs in humans. Located near the base of the neck, human larynx is essential for voice production, breathing, and swallowing, and also functions to protect the lungs from foreign objects and helps forming a rigid thorax that supports the upper body during activities like lifting and pushing (Aronson, 1985). The role of the larynx in speech production from a comparative and phylogenetic perspective can be considered as a recent event. Voice production and communication, however, seems to be another story. Warning, territorial, and release calls in frogs (200 million years ago) can be given as examples to this.

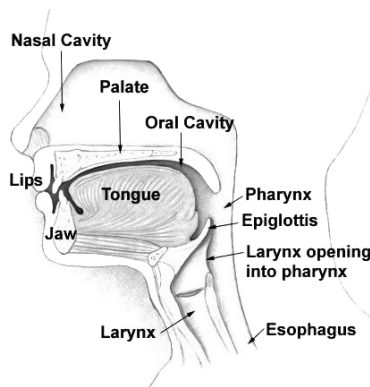


Figure 2.1: Overview of the basic components involved in voice production and the surrounding organs. Public domain image from Wikimedia Commons.

2.2 Evolutionary perspectives on voice cognition

Evolutionary psychology provides a perspective on the origins of human cognition and on the functional components that gave organisms a selective advantage

when dealing with the problems in ancestral situations. A knowledge of our phylogenetic inheritance can deepen the proximate explanations offered by cognitive and social psychology, and neuroscience. In this subsection, I will apply the analysis of adaptive problems to the study of voice cognition to provide evolutionary perspectives to these questions: What is the function of voices? What are their origins and how did they assume their distinctive role in human cognition? In particular, I will cover what adaptive problems voice may be evolutionarily designed to solve, and explain the ontogenetic and phylogenetic origins of voice cognition².

2.2.1 Adaptive problems related to voice

From an evolutionary standpoint, the survival of humans in ancestral situations was critically dependent on quickly and accurately responding to various signals of threat (Buss, 2012), or in Darwin's (1859) terms to the 'hostile forces of nature', even in the absence of visual cues (e.g., in the darkness). Rapid auditory recognition abilities and identifying the source of a voice and sound directly address the adaptive problems of, for example, detecting predators or prey, thunder, etc. Voice recognition, as an evolved psychological mechanism, functions also to detect hostile conspecifics (e.g., separating foe from friend).

Another adaptive value of voice recognition is related to kin recognition, which functions in solving problems of parenting and aiding genetic relatives. These can include facilitating reunion between the parent and the offspring, ensuring parental investment (i.e., investing in the care of one's own children), and kin altruism. All of which, according to the inclusive fitness theory (Hamilton, 1964), are essential in promoting inclusive fitness (i.e., the fitness of the individual plus the reproductive success of the genes related to oneself).

Finally, voice cognition is clearly critical for social communication. Social communication functions to solve problems of survival and group living (e.g., resolving conflict, social bonding, and forming alliances). Indeed, vocal expressions (including language, laughter, and singing) were argued as a form of 'social grooming' (Dunbar, 2017). A highly prominent use of voice in communication is the social interaction between caregiver and child, which will be covered in the next subsection. Since voices convey a myriad of information about an individual (e.g., age, appearance, health status, emotional status, etc.) evolved mechanisms of voice cognition are useful in addressing problems of mating as well (evaluating reproductive fitness of a potential mate, e.g., selecting and attracting a mate who is healthy, able to invest, is compatible etc.).

²A warning might be useful here. Providing a full list of the adaptive problems that gave rise to human voice cognition mechanisms is simply impossible. I followed the broad classes of adaptive problems from the modern evolutionary theory to give insights to the questions posed above.

2.2.2 Ontogenetic origins of voice cognition

Voices, like faces, are of great significance in ontogenetic and phylogenetic survival (Kreiman & Sidtis, 2011). Voice processing abilities are found very early in human life. In fact, there is evidence suggesting that voice recognition begins before birth. Using various methods, several developmental studies have provided evidence for the early perceptual preferences for voices in neonates and in fetuses. For example, DeCasper & Fifer (1980) revealed how newborns change their duration of sucking behaviour to hear their mother's voice, suggesting neonates' preference for own mother's voice over another female voice. Using the same method, Fifer & Moon (1994) also observed a neonatal preference for maternal voice in response to conversational prosody. They also found a consistent change in neonates' and in premature fetus' (at 26-34 weeks of gestation) heart rate in response to vowels produced by a male voice, indicating an orienting response to voices in general. With heart rate measures, Kisilevsky et al. (2003) showed evidence for the ability of the fetus to recognize human voices that they experienced in utero. In particular, their findings indicated an increased heart rate in response to mother's voice and decreased in response to a strangers' voice.

A related question is whether such a preference for human voices in humans is innate or experience-dependent. Vouloumanos et al. (2010) compared human neonates' (37.5 weeks of gestation) and infants' (3 months old) preference for voices of humans and non-human primates. They found that while the neonates did not have a preference for human voice (nonsense words) over the calls of rhesus monkey, the infants showed a bias for the human voice material over the rhesus calls. This 'auditory tuning to conspecifics' by 3 months is significantly earlier than its visual counterpart (appearing at 9 months). The authors argue that this rapid auditory tuning may reflect a special status of speech.

In a more recent study by Webb et al. (2015) cranial ultrasonography was used to investigate whether exposure to maternal sounds would elicit structural alterations in the auditory cortex for infants born prematurely (25 to 35 weeks gestational age). Strikingly, their results revealed that the auditory cortex was larger in neonates exposed to maternal sounds (i.e., audio recordings of the mother's voice and heartbeats) than in the ones in the control group that was exposed to environmental sounds (i.e., hospital noise). These findings indicate an experience-dependent auditory plasticity in newborns.

Collectively, these findings are suggestive of the importance and the early developmental onset of specialized voice processing mechanisms. It is clear that these findings also have important implications for both voice and speech perception studies.

2.2.3 Voice cognition among nonhuman species

Voice processing and production abilities are not unique to humans. In most cases, vocal communication takes place between conspecifics and in certain situations (e.g., predator-prey settings) heterospecific vocalizations are also used

to behave adaptively in the context (Andics & Faragó, 2018). A particularly active area of research within several branches of biology is the study of songs (i.e., birdsongs). A large body of research underlines the importance of birdsongs in mating behaviour and in repelling rivals (Collins, 2004). Several bird species can skillfully discriminate conspecific and heterospecific songs, but they are generally better at discriminating among various songs of conspecifics than the songs of heterospecifics (Knudsen & Gentner, 2010). In insects, the studies have revealed that neural tuning for temporal pattern recognition of courtship songs guides response in mating behaviour (Baker et al., 2019). Many other nonhuman animals, including (but not limited to) primates, bats, frogs, wolves, and penguins use species-specific vocalizations for individual recognition and discrimination (Kreiman & Sidtis, 2011).

Due to having the closest evolutionary relationship to humans, the studies with nonhuman primates are especially attractive to indicate the evolutionary roots of voice and speech processing. Most relevant to this thesis, these studies are key to answering an essential question: Are conspecific voices special in the way they are processed compared to other sounds (Belin, 2006)? A recent review by Bodin & Belin (2020) offers an extensive overview of the evidence for the conspecific voice perception in primate brains. Based on this evidence, they suggest a voice patch system in primates, that is, a network of interconnected voice-sensitive areas which can serve as a common template for the cerebral basis of voice perception.

Most clear behavioural and neural evidence comes from the studies on rhesus macaques, the primate line that have a common ancestor with modern humans roughly 25 million years ago. Behavioural studies reveal that the macaques are able to use identity information from conspecific vocalizations (Bodin & Belin, 2020). For example, behavioural evidence exists for voice recognition of both individuals and kin among free-ranging macaques (Rendall et al., 1996). Macaques also perceive changes in formant frequencies in conspecific vocalizations (Fitch & Fritz, 2006), possibly to make predictions about vocalizer's body size (Fitch, 1997).

Imaging studies have shown preferential responses to conspecific voices, suggesting functional similarities between human and monkey brain for voice cognition. For instance, Petkov et al. (2008) observed voice specific regions in the macaque brain using functional magnetic resonance imaging (fMRI). In particular, the anterior superior temporal plane, a higher-level auditory region, exhibited stronger responses to conspecific vocalizations than to various other acoustic stimuli (including other animal vocalizations, environmental sounds, and acoustic controls sharing the spectrum and duration with the species-specific vocalizations). In addition, they observed a preferential response for familiar conspecific vocalizations (i.e., vocalizations of the conspecifics belonging to the same colony as the monkey). In a third experiment, by systematically varying the conspecific individual and the vocalization (i.e., coo or grunt), they also found that the same vocalization from different individuals revealed the highest sensitivity and that the right anterior region was found to be more sensitive to distinguishing the identity of individuals. Furthermore, Perrodin et al. (2011)

identified specialized ‘voice cells’, a population of neurons with strong preference for conspecific voices compared to environmental sounds and heterospecific voices, located in anterior temporal plane in macaques. This provides direct evidence for the neural underpinnings of the voice-selective areas found in earlier fMRI studies.

In sum, the evidence indicates that the voice of conspecifics is special and may have dedicated networks in each species’ brain. Together, these studies point to the ancient evolutionary roots of the neural architecture involved in voice cognition.

2.3 Behavioural and neuroscientific perspectives on human voice cognition

In this subsection, I address behavioural, electrophysiological, and neuroimaging studies on human voice cognition. These contribute to understanding why human voices may be unique (for most humans), as well as the neural mechanisms underlying voice cognition in humans. Although they also provide important insights into voice cognition, studies on the topics of emotional voice recognition and familiar voice recognition, and neuropsychological cases relating to voice-identity deficits are excluded, since the issues discussed in the papers in this thesis do not concern any of these topics.

2.3.1 Human voice areas in the brain

Along with the evidence from primates, the presence of voice-selective areas in the human brain is also provided in the literature. Using fMRI, Belin et al. (2000) showed areas in human brain that are sensitive and strongly selective to human voices (including speech and non-speech stimuli). Greater neuronal activation in these areas mostly distributed along the STS were demonstrated for voices compared to various non-voice sounds and acoustic control stimuli in three experiments; ruling out that this activation was in response to a specific acoustic component in the voices or to sounds with human origin (e.g., hand clap).

Complementary evidence for the existence of the so-called ‘temporal voice areas’ in the STS that preferentially respond to human voices was provided by several other studies using different stimuli (e.g., Belin et al. 2002; Fecteau et al. 2004; Kriegstein & Giraud 2004; Pernet et al. 2015; Agus et al. 2017; Whitehead & Armony 2018). Based on the finding that non-speech stimuli induced larger activation than frequency-scrambled control stimuli only in the right anterior STS, an exclusive involvement of this part in perception of voice without linguistic content was suggested (Belin et al., 2002). A comparison of the brain responses in the STS elicited by human and animal vocalizations, and musical instruments revealed a greater bilateral response in the anterior STS for human voices (speech and non-speech) compared to both animal and non-vocal (musical instruments and environment) sounds (Fecteau et al., 2004).

One might argue that this preferential response to human voices in the STS might be due to familiarity with the stimuli. In the study by Fecteau et al. (2004), cat vocalizations, a stimulus that human subjects are highly familiar with, in contrast with human voices also indicated bilateral STS activation, as in the contrast of human and mixed animal vocalization.

While STS responds more to human voices, musical stimuli seem to induce greater responses in superior temporal gyrus, especially in planum polare (Whitehead & Armony, 2018). Aiming to provide a comprehensive picture of the neural responses to music that is not only limited to musical instruments, Whitehead & Armony (2018), showed that *a capella* singing, an intermediate condition between these two types of stimuli, elicited a neural overlap of these areas. Furthermore, a recent study provided evidence for song-selective neural populations in the human brain, distinctive from neural responses to instrumental music and speech (Norman-Haignere et al., 2022).

2.3.2 Are human voices processed more quickly than other auditory stimuli?

One way to infer whether human voices are processed quicker than other stimuli is through comparing response times (RTs). Given that human information processing in the brain is highly structured, the pathways in this structure will have varying time courses. The idea is that these should also be reflected through RTs (Luce, 1986). Agus et al. (2012) conducted behavioural experiments comparing recognition of voices and musical instruments in a go/no-go task (where listeners were asked to respond by releasing a button as fast as possible to target sounds and withholding response to distracters). By measuring RTs during this task, they observed faster responses for human voice targets compared to targets belonging to percussion and string instruments, and for voice-instrument chimeras³ among musical instrument and instrumental chimera distracters.

More recently, Isnard et al. (2019) used the fastest presentation rate that allows for accurate recognition as a measure of processing time. In this behavioural study, voices could still be recognized among instrumental sound distracters at a very rapid rate (i.e., 30 Hz), and the fastest processing times were obtained for human voices. In addition, overall faster RTs were observed when a voice target was presented within a rapid sequence of instrumental sound distracters compared to instrument sound target among voice distracters (in all except the slowest presentation rate of 5.3 Hz, where the RTs were similar).

In concert with the behavioural findings of faster processing of human voices over other auditory stimuli, evidence from electrophysiological studies supports the idea that human voices elicit rapid cortical responses as compared to other sounds. These studies typically employ a detection task embedded in an oddball paradigm, where the participants press a button each time they hear the target while ignoring the distracter sounds, and event-related potentials (ERPs) elicited

³These were created using spectral and temporal features of the voice and instrument stimuli by preserving one feature from each category.

by the distracters belonging to various categories are compared. For example, Levy et al. (2001) found that human voices elicited a stronger positive component peaking at 320 ms post-stimulus onset compared to string, wind, and brass instruments, suggesting a "voice-sensitive response". In the light of the findings from a follow-up study, the same group later proposed that this response might be associated with attentional capture by certain voice and stimuli perceived as voice-like (i.e., strings) (Levy et al., 2003). Upon closer examinations of the ERP responses to string instruments, Levy et al. (2003) also showed that the largest positivity among these was evoked by the cello tones.

Using the same stimuli as Levy et al. (2001), in a magnetoencephalography (MEG) study Gunji et al. (2003) compared evoked magnetic responses to human voices and string instruments. The task involved the participants watching a silent film while hearing (but instructed not to pay attention to) the sounds. In contrast to the findings presented above, the magnetic counterpart of the voice sensitive response was not observed in this study. The authors interpreted this finding as potentially resulting from the radial orientation of the neuronal electromagnetic sources or alternatively, from watching a film in this study, as compared to Levy et al. (2001)'s study where there was no viewing condition. In a later study, larger ERPs in response to voices were found much earlier when compared to bird songs and environmental sounds, which included natural sounds, musical instrument sounds, and mechanical sounds Charest et al. (2009). In particular, this "fronto-temporal positivity to voices" started to emerge as early as 164 ms after stimulus onset and peaked at around 200 ms, suggesting a rapid discrimination of voices among non-voice stimuli. This processing speed of the brain for differentiation of voice and non-voice stimuli is very comparable to the speed of differentiating face and non-face stimuli, which takes about 170 ms (Belin, 2019).

To put these numbers into a slightly wider context, I would like to refer to a study by Murray et al. (2006) where stimuli were categorised more broadly. Again, using the oddball paradigm, Murray et al. (2006) observed ERP responses differentiating sounds of living (i.e., animal sounds, human vocalizations) and human-made auditory objects (i.e., various musical instrument sounds and environmental sounds such as bike bell, police siren, glass shattering etc.) as early as 70 ms post-stimulus onset. Furthermore, comparing ERPs elicited by the same sounds serving either as distracters or as targets, they observed earliest differential responses at 100 ms after stimulus onset. The authors proposed this as an upper limit on the speed of auditory discrimination in the brain.

2.3.3 Are human voices detected, categorized, and recognized better than other auditory stimuli?

Quite surprisingly, behavioural research comparing processing of human voices in contrast with other stimuli is rather scarce. Despite the contextual differences resulting from different study designs, behavioural studies (e.g., Agus et al. 2012; Bigand et al. 2011; Suied et al. 2014; Isnard et al. 2019) have generally shown a

behavioural advantage in detection, categorization, and recognition of human voices as compared to musical instruments and environmental sounds.

In addition to RTs, Agus et al. (2012) compared task accuracy through averaging false alarm rates (i.e., responding incorrectly as target to a distracter) for different sound categories. False alarm rates were lowest for human voices (4%) which was lower than the strings (12%) at 0.05 significance threshold. Human voices were categorized with less false alarms than the strings also when the target-distracter similarities were equated. Lowest false alarm rates for human voices remained also when compared to chimeras (created with various dual combinations of strings, percussions, and human voices) among both natural and chimera distracters.

Comparing the accuracy for categorizing spoken voices, instrumental classical music, and environmental sound stimuli normalized using two different amplitude normalization procedures, Bigand et al. (2011) found that a voice advantage (i.e., higher accuracy for the categorization of human voices than other stimuli) was observed for sounds as brief as 30 ms, but only when a root mean square (RMS) normalization was applied. Theoretically, evidence for lack of an effect of amplitude normalization method would have suggested that the auditory processing is modular (i.e., automatic and fast processing of human voices). Since a difference was observed, these results were interpreted to reflect a general categorization process, where a human voice advantage results from the variability of the spectral envelope between and within sound categories (Bigand et al., 2011).

In the experiments by Suied et al. (2014), listeners were asked to report whether they heard a target category among other sounds. Human voices (i.e., sung vowels, as in Agus et al. 2012) presented as targets among other human voices and among musical instruments as distracters were recognized better than when instrumental sounds served as targets (among other instruments and among human voices), even in the briefest stimulus durations. Using similar stimuli, in various stimulus presentation rates, Isnard et al. (2019) also reported a better recognition for voice targets among instruments than the reverse situation.

The choice of stimuli in these studies were criticised in the sense that that sung vowels and isolated musical tones are weakly representative of voices and musical sounds (Bigand et al., 2011). Studies using more ecological stimuli seem to corroborate previous studies where a human voice advantage was observed. For example, comparing human adults' memory for vocally and instrumentally presented melodies using an old-new recognition task, Weiss et al. (2012) revealed that vocal melodies were remembered better than instrumental melodies.

2.3.4 Similarities and differences between voice and face cognition

Recent behavioural, electrophysiological, functional imaging, and neuropsychological studies on humans have provided converging evidence for the significance of human voice processing. The available evidence is largely paralleling those related to human face processing and in support of the idea of similar and

2. Voice cognition

interacting functional structures for processing these two stimuli, as suggested by Belin and colleagues (2011). Indeed, pioneering models of face recognition (Bruce & Young, 1986; Burton & Bruce, 1992) have integrated voice recognition. Inspired by these models, a similar functional architecture has been proposed for voice perception (Belin et al., 2004). Related to these, the metaphor of voice as an ‘auditory face’ mentioned earlier in this chapter encapsulates the idea behind the similarities between face and voice processing. A summary of these similarities is presented by Yovel & Belin (2013), which can be found in Table 2.1.

Table 2.1: Summary of some of the evidence for the similarities between face and voice processing. References for the findings presented in the table can be obtained from the original article. This article was published in Trends in Cognitive Sciences, Vol. 17, No. 6, Galit Yovel and Pascal Belin, A unified coding strategy for processing faces and voices, 263-271. Copyright Elsevier (2013). Table reproduced and reprinted with permission from Elsevier (2022).

	Face	Voice
Neural selectivity		
Human		
<i>Electrophysiology</i>	N170/M170	FTPV
<i>Functional MRI</i>	Face areas in the lateral occipital, mid fusiform and STS	Voice areas in the STS
<i>Hemispheric asymmetry</i>	Right hemisphere	Right hemisphere voice-selectivity (left hemisphere for speech)
<i>Effects of TMS</i>	TMS over the OFA selectively impairs performance for faces and selectively increases the face N170	TMS over the TVA disrupts voice detection
Monkey		
<i>Electrophysiology</i>	Face-selective cells	Voice-selective cells
<i>Functional MRI</i>	Face-selective brain areas	Voice-selective brain areas
Selective recognition deficits		
	Developmental and acquired prosopagnosia	Developmental and acquired phonagnosia
Perceptual Coding		
<i>Norm-based coding</i>	Relative to an averaged face	Relative to an averaged voice
<i>Distinctiveness effect</i>	Better recognition for distinctive faces	Better recognition for distinctive voices
<i>Perceptual aftereffects to anti-faces/voices</i>	Largest for matched vs non-matched anti-faces	Largest for matched vs non-matched anti-voices
<i>Attractiveness</i>	Averaged face is more attractive	Averaged voice is more attractive
Development and experience		
<i>Early preference</i>	Preference for upright faces 24 hours after birth	Fetuses and young infants discriminate voices from other auditory stimuli
<i>Neural correlates</i>	Face-selective ERPs appear at three-six months	Voice areas emerge between three and seven months
<i>Perceptual narrowing</i>	Broad abilities for cross species face recognition at four-six months are tuned by experience in eight-ten-month-old infants	Broad abilities for phoneme discrimination at four-six months are tuned by experience in eight-ten-month-old infants
<i>Effects of experience in adulthood</i>	Other race effect	Language familiarity effect and own-race bias

This does not mean, however, that there are no differences between face and voice cognition. Some of the differences between face and voice perception are mentioned by Kreiman & Sidtis (2011). These include, for example, that voices exist in the temporal domain, while faces exist in the spatial domain. According to Kreiman & Sidtis (2011), voices happen at a certain physical distance, while

faces occur immediately⁴. These authors also suggest that energy is produced from voices, but less so from faces.⁵ Unlike faces, voices can be heard in utero and are recognized developmentally much earlier than faces. Hanley (2008, p.135) states: “Voices may express mood, affect, and emotions even more clearly than faces.”.

In a recent paper by Young et al. (2020), it was argued that the theoretical models of voice perception should take the consequences of the differences between face and voice perception into account. For example, they suggest that the integration between face and voice signals is more evident for emotion recognition and less so for identity recognition. Providing a potential background to this, they point out, among other things, the differences in the temporal demands of auditory and visual signals. That is, while cues related to identity recognition (e.g., age, gender) would have high temporal demands as these remain relatively stable once established, cues related to emotion recognition would require constant monitoring as these would show moment to moment variability.

To conclude, studying voice processing based on its similarities to and differences from face processing in various contexts has important theoretical implications, as it can help to better understand these processes and the potential cross-talk between them.

As we will discover in the next chapters of this thesis, I draw inspiration mostly from the similarities between face and voice processing, but there is certainly room for research focusing on both the similarities and the differences between them.

⁴Kreiman & Sidtis (2011) acknowledge here that this difference is changing due to technological advancements.

⁵One can also argue that voices produce a mechanical energy, but faces are surfaces that reflect light energy.

Chapter 3

Temporal limits of attention

In this chapter, I will make introductory remarks on the temporal constraints of human information processing, and discuss how this topic is situated in the auditory modality. Next, I will visit some of the theories underlying temporal limits of attention and also showcase, based on previous studies, certain situations where these constraints appear to be less likely to occur. Finally, I will explore the topics of musicality and musicianship, and connect them to the topic of temporal selective auditory attention.

3.1 Introduction

The majority of cognitive functions are based in time (Cohen, 2014). Attentional mechanisms are no exception to this rule. Paraphrasing Shapiro (2001) who regards the study of temporal attention central to several fundamental issues, the following questions can be placed at the heart of these issues: Do humans process information continually or does this ability show intermittency? If the latter, then what are the limits of the human cognitive system? Finally, is it possible to have control over these limits?

As stated in Chapter 1, this thesis investigates whether these temporal limits, particularly in the auditory modality, can be modulated through various experimental manipulations. Before we can explore ways to modulate the limits of attention, we need to first understand what evidence there is to suggest the existence of these limits and how to study them.

Attentional blink (AB), psychological refractory period, repetition blindness/deafness, inattention blindness/deafness and change blindness/deafness are a family of cognitive phenomena generally known to reflect the so-called limits (or constraints) of the human attentional system and are heavily studied in the field of cognitive psychology. While this thesis is primarily concerned with the AB phenomenon, each of these phenomena connects to their own rich literature that extends beyond this thesis. It may, however, still be useful to provide definitions of the AB alongside its cousins¹, which can be found in Table 3.1. Although the paradigms used in the study of most of these attentional phenomena were, to a large extent, designed to investigate the selection and processing of objects and features in visual search contexts (Simons & Chabris, 1999), not all paradigms originated in the visual domain. For example, the

¹I use this word mainly to indicate that these are phenomena are similar and to some extent related, but it should not be assumed that they are closely related to attentional blink nor should it be inferred that the underlying cognitive processes for these phenomena are identical. For example, previous research showed repetition blindness to be highly dissociable from the attentional blink effect (Chun, 1997).

3. Temporal limits of attention

paradigm used to study inattentional blindness was devised as a visual analogue of the dichotic listening task (i.e., simultaneously listening to different auditory information presented to each ear), first introduced by Cherry (1953).

Table 3.1: Attentional blink and its cousins

Phenomenon	Also known as in the auditory domain	Definition
Attentional blink (Raymond et al., 1992)	-	Failure in reporting a second target after correctly identifying the first target when the targets closely follow one another within a rapid serial presentation of stimuli.
Psychological refractory period (Welford, 1952)	-	Slower response time for the second task when participants are presented with two stimuli in close temporal proximity and asked to respond to these stimuli with two separate tasks in a speeded fashion.
Repetition blindness (Kanwisher, 1987)	Repetition deafness (Miller & MacKay, 1994)	Failure in reporting the second occurrence of a repeated target in a rapid serial presentation of stimuli.
Inattentional blindness (Mack & Rock, 1998)	Inattentional deafness (Macdonald & Lavie, 2011)	Failure to notice an unexpected stimulus while performing an attentionally demanding task.
Change blindness (Rensink et al., 1997)	Change deafness (Vitevitch, 2003)	Failure in noticing large changes in the scene or in the object during a brief disruption.

In what follows, I will describe the classic AB paradigm and the study which led to the notion of dual-task interference, which is considered as the foundation of the AB phenomenon.

3.2 Dual-task costs in rapid serial presentation paradigm

Anecdotally, people rarely notice difficulty in performing two tasks simultaneously. However, laboratory studies suggest that carrying out two (or more) activities concurrently often hampers performance significantly, even in situations where these activities are not “physically incompatible” and/or “intellectually demanding” (Pashler, 1994, p. 220). After more than a century of research on dual-task performance, we know that dual-task can worsen the performance but we still do not know precisely why this is (Hommel, 2020). Two main classes of theoretical frameworks offer possible answers to the why: the resource account (e.g., Kahneman 1973) and the processing stage account (e.g., Sternberg 1969). The resource account’s answer is based upon the idea that an increased number of processes would likely lead to an overload of the cognitive system, as its resources are capacity-limited. Conversely, processing stage models suggest that human information processing runs through operating stages, among which response selection is primary and does not allow for parallel (i.e., simultaneous) processing.

Much like many other dual-task situations, reporting of dual-targets have been shown to interfere with one another (Broadbent & Broadbent, 1987; Duncan, 1980). In their classic paper, Broadbent & Broadbent (1987) undertook research which explored the limits of interference between reporting two targets amongst rapidly presented word lists. This involved presenting subjects with a rapid serial visual presentation (RSVP) stream with a series of lowercase words, which were presented sequentially at the same location. Each stream included two pre-defined targets separated at various intervals (by intervening non-targets). Their findings showed that when the targets were temporally close to one another, correct responses were given to either target but not to both. When T1 was correctly identified, the performance on T2 identification deteriorated, but started to recover again at 720 ms stimulus onset asynchrony (SOA). According to Broadbent & Broadbent (1987) the presence of a first target disrupted the identification of T2 because of the higher demands of the identification process (as opposed to detection, which can be done in parallel). This study is considered as the first demonstration of an AB effect (Dux & Marois, 2009).

3.3 Attentional blink phenomenon

In analogy to an overt blink of an eye, Raymond et al. (1992) coined the term *attentional blink* because of the brief temporal nature of this phenomenon, lasting for about half a second. AB is a cognitive phenomenon describing that when two targets (T1 and T2) are presented within a rapid presentation stream, the reporting of the second target (T2) is impaired when T2 occurs shortly after T1 (Raymond et al., 1992; Shapiro et al., 1997b). Simply put, the phenomenon suggests that people typically fail at reporting the T2 when it closely follows T1, “as if the perceptual and attentional mechanisms blink” right after identifying T1 (Raymond et al., 1992, p. 851). This phenomenon has played a key role in delineating the limits of attention and memory processes, as well as of consciously processing stimuli distributed over time (Dux & Marois, 2009; Martens & Wyble, 2010).

In the seminal paper by Raymond et al. (1992), participants were presented with a stream of letters displayed centrally on a screen with a gray background. Among the non-targets (shown in black), T1 was shown in white. T2 was the letter X and only shown in half of the trials. Critically, the serial positions where T2 could appear were experimentally manipulated. The participants first had to identify T1 letter and then indicate whether T2 was present or absent (Experiment 2, see Raymond et al. 1992 for further methodological details). Two important findings came out of this paper. First, a dramatic deficit in detecting T2 was observed for 180 to 450 ms SOA. This finding clearly showed that the deficit observed by Broadbent & Broadbent (1987) with an identification task could also be observed in a simple detection task. Second, this deficit was not present in the control condition where participants were to ignore the T1, suggesting that the AB reflects not a sensory but an attentional process. Following this iconic paper, a substantial number of experiments have

3. Temporal limits of attention

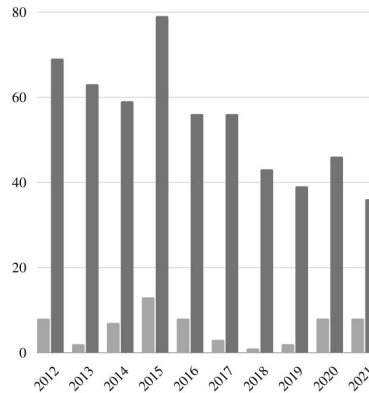


Figure 3.1: Number of publications appearing on Web of Science between 2012 and 2021. Search results for "Auditory Attentional Blink" (shown in the figure with light grey) and "Visual Attentional Blink" (shown in the figure with dark grey) based on "Topic" (i.e., the article title, abstract, author keywords, and keywords plus). The search included the research articles only.

successfully replicated the AB phenomenon. In addition to the use of words, numbers and letters, some of these studies have also included a variety of other stimuli (e.g., symbols, shapes, images of scenes or objects, sounds). Central to this thesis is the auditory variant of the AB task ².

3.3.1 The auditory attentional blink

Investigations of AB in the auditory modality have thus far shown evidence for this phenomenon using simple tones (e.g., Shen & Mondor 2006; Goddard & Slawinski 1999; Shen & Alain 2010), spoken letters and digits (e.g., Arnell & Larson 2002; Arnell & Jenkins 2004; Martens et al. 2015), and spoken syllables (e.g., Tremblay et al. 2005). Despite an overwhelming interest in the visual AB phenomenon, the investigations in the auditory modality (as well as in other sensory modalities) have been limited. Figure 3.1 shows the number of publications in the last decade comparing the search results for auditory AB and visual AB. The auditory AB, nonetheless, roused debates that lasted for a long time.

Modality debates

The majority of the earlier auditory AB experiments were conducted primarily to resolve whether attention is limited by a common, central attentional system

²As this has been rather understudied, findings from the visual modality (whenever relevant) have also been discussed in this chapter.

or by independent attentional resources where each modality is restricted by its own limitations. These experiments typically contrasted results obtained from each single modality (auditory versus visual) and/or from mixed-modality tasks (i.e., audio-visual).

First evidence for an auditory AB came from Duncan et al. (1997) who investigated the time-course of the dual-target interference both between and within auditory and visual sensory modalities, using the methodology that later became known as attentional dwell time. The difference in this methodology is that the target events are spatially separated. The auditory experiment involved presentation of non-target spoken syllables among which target syllables were embedded in two concurrent sound streams (one stream with low-pitch and the other with high-pitch voice). Their findings showed similar results in single modality auditory and visual tasks. When the participants were to identify both targets, the target identification accuracy was substantially lower when T2 followed T1 in the other stream within a few hundred milliseconds. Whereas in the mixed-modality experiment (i.e., where one stream was auditory the other was visual), dual-task interference was no longer observed. The authors concluded that attending to concurrent targets did not show restriction between modalities, suggesting modality-specific attentional restrictions.

Aiming to investigate whether the attentional deficits observed in the visual AB phenomenon would also be observed in the auditory modality and in the cross-modal context, Potter et al. (1998) conducted six AB experiments using written and/or spoken letters and digits. In these studies, no statistically significant lag effect was observed in either auditory sequences or in the mixed-modality T1 and T2s. Potter et al. (1998, p. 981-982) concluded that “with auditory stimuli there is no evidence for a lag effect” and “auditory attention does not blink”³. They suggested that this may be due to the larger capacity of echoic memory, which allows for handling temporal events over an extended period in the auditory modality. They argued that the presence of a task-set switch (i.e., the cost of changing from one sort of target in T1 task to another sort in T2 task, Potter et al. 1998) in most of the experimental paradigms accounts for the observations of auditory and cross-modal AB effects.

Goddard & Slawinski (1999) argued that spoken words and letters (due to their visual representations) might not fully reflect a purely auditory effect. Instead, they compared AB in auditory and visual modalities using pure tones and lines. The same individuals performed both the frequency identification and line orientation tasks within the AB paradigm. Their results revealed an AB effect in both modalities, lasting for the first four inter-target lags (SOAs = 90, 180, 270, and 360 ms). Interestingly, in this study, the magnitude of the effect was markedly larger (around 2.5 times) in the auditory modality than in the visual modality. Moreover, individual performance on the auditory AB task did not show statistically significant correlation with the visual AB task

³The statistical basis for these claims can be questioned, since these conclusions were based on non-significant results (exact *p*-values for these results were also not reported) and as the “absence of evidence is not evidence of absence”.

3. Temporal limits of attention

performance. The authors interpreted these findings to suggest modality-specific attentional mechanisms governing the AB effect.

To resolve the divergence across the findings of the between- and within-modality audio-visual AB investigations, Arnell & Jolicoeur (1999) conducted four experiments, each designed so that the modalities of both targets were fully crossed. Their stimuli included numbers and letters (spoken or visually presented), except in Experiment 3 where pure tones were used. Together, the results from these experiments showed both auditory and cross modal AB effects, suggesting that the AB phenomenon is not unique to the visual domain. Importantly, the size of the AB effect was comparable in both modalities, independent of whether the two targets belonged to same or different modalities (Experiment 1). This led the authors to suggest a central attentional framework underlying the AB, with minimal or no modality effects. Cross-modal AB effects were also found when pure tones were used as the auditory stimuli (Experiment 3). These researchers suggested a systematic link between the stimulus presentation rate and the AB. Testing this link, they found evidence for a modulation of the AB magnitude by the stimulus presentation rate in both the purely auditory and the purely visual AB task, with an attenuated AB magnitude as presentation rates were slowed down (Experiment 4).

More recently, the time course of auditory AB was examined via ERP recordings (Shen & Alain, 2010). According to Shen & Alain (2010) if there would be similar neural signatures for auditory and visual AB, this would provide support for common underlying mechanisms in both modalities. In addition to Lag (0, 1, 7), Shen & Alain (2010) also manipulated the presentation rate as a way to modulate the size of auditory AB through varying the SOA (90 ms, 120 ms, 150 ms) between each tone (lasting for 30 ms) in the stream. They found that the size of auditory AB increased with increasing presentation rates. In addition, larger but delayed P3b amplitude was observed in slower presentation rates compared to the faster rates in this study. Due to the similar electrophysiological results compared to the earlier visual AB studies (e.g., Vogel & Luck 2002), this study showed evidence for a common ERP signature elicited by T2 during AB.

A debate within a debate Directly testing Potters' claim regarding task-switch, Arnell & Larson (2002) examined the auditory (as well as cross-modal and visual) AB effect in a design that did not allow for a task-set switch reconfiguration, regardless of whether it may be stimulus-driven or related to participants' readiness. AB effects were observed in both the auditory and visual modalities (cross-modal AB was found only when T2 was visual). Similarly, Soto-Faraco & Spence (2002) has brought into attention Duncan's (1997) findings where auditory AB was found without a task-switch, conflicting with Potter's claim. However, as described earlier, in this study targets were presented in two concurrent streams. Therefore Soto-Faraco & Spence (2002) tested auditory AB that did not include a task-switch or a spatial aspect (i.e., targets appearing in a single auditory stream). In this study, the auditory stimuli consisted of spoken letters and digits, each lasting for 85 ms. The task involved participants to type

on a keyboard the digits presented in the stream irrespective of the modality and order. They found a significant auditory AB, as shown by increased T2|T1 accuracy with lag (increase from Lag 2-3 through Lags 5-7). Thus, contrasting Potters' explanation for the auditory AB results in the literature, this study showed that auditory AB can be present also when there are no task-switch costs. As mentioned above, another basis for the claims of AB as a purely visual phenomenon was the idea that auditory stimuli may be more resistant to masking from subsequent events in the stream than visual stimuli (i.e., Potter et al. 1998; Chun & Potter 2001). However, Vachon & Tremblay (2005) provided evidence for functional similarities across auditory and visual modalities by showing that, just as in vision, masking is necessary for auditory AB and is effective even when delayed.

3.3.2 Theories of attentional blink

Many theories (as shown in Table 3.2) have been proposed to account for the AB effect (for comprehensive reviews please see Dux & Marois 2009; Martens & Wyble 2010), but no theory thus far can fully account for the findings associated with this complex phenomenon. Presumably due to the common use of visual stimuli and the assumption that AB arises from early processing limitations (Arnell & Larson, 2002), most theories were developed and tested largely within the visual domain.

Initially, the theoretical frameworks of AB generally assumed that AB results from a capacity limitation (i.e., that people do not recognize more than one stimulus at a time) or processing bottlenecks that temporarily impair T2 detection (e.g., Chun & Potter 1995; Jolicoeur 1998; Dux & Harris 2007). The two-stage model has been a particularly influential theory, which postulates that every stimulus presented in RSVP goes through two processing stages: in the first stage their conceptual representations get activated, allowing them to be rapidly detected. During Stage 1, however, these representations can easily be forgotten or overwritten by a subsequent stimulus. For a target stimulus to be successfully reported, it needs to be further processed and consolidated in Stage 2. According to this model, AB occurs because of the capacity limitations of Stage 2, that is, while Stage 2 is still pre-occupied with T1 (under short lags where there is a processing bottleneck for T1), T2 cannot be consolidated into WM (Chun & Potter, 1995; Jolicoeur, 1998).

Due to lack of space, I will not be able to cover all theories ⁴, but because of its relationship to one of the methodologies used in this thesis, next, I will describe the neurocomputational model of AB (Nieuwenhuis et al., 2005), which is based on the activity of the locus coeruleus-noradrenaline (LC-NA) system.

⁴This is already done in the excellent review papers by Dux & Marois (2009), covering the full extent of both the formal and informal AB theories, and by Martens & Wyble (2010), which brings together the past, present and the future of AB research and theories.

3. Temporal limits of attention

Table 3.2: Overview of the theories of attentional blink. Table produced based on the review paper by Dux & Marois (2009). References for the theories presented in the table can be obtained from the original article.

Theoretical account	Reference
Informal theories of AB	
Inhibition hypothesis	Raymond, Shapiro, & Arnell (1992)
Revisited inhibition hypothesis	Olivers, van der Stigchel, & Hulleman (2007)
Interference theory	Shapiro, Raymond & Arnell (1994)
Two-stage model	Chun & Potter (1995)
Attentional dwell hypothesis	Ward, Duncan, & Shapiro (1996)
Central interference theory	Jolicoeur (1998, 1999)
Two-stage competition model	Potter, Staub, & O'Connor (2002)
Multiple channel hypothesis	Awh et al. (2004)
Extension of two-stage bottleneck theory	Dux & Harris (2007)
Temporary loss of control hypothesis	Di Lollo, Kawahara, Ghorashi, & Enns (2005)
Delayed attentional reengagement account	Nieuwenstein (2006); Nieuwenstein, Chun, van der Lubbe, & Hooge (2005); Nieuwenstein & Potter (2006); Nieuwenstein, Potter, & Theeuwes (2009)
Hybrid models of AB	
	Vogel, Luck, & Shapiro (2008) Kawahara, Enns, & Di Lollo (2006)
Formal theories of AB	
Gated auto-associator model	Chartier, Cousineau, & Charbonneau (2004)
Corollary discharge of attention movement model	Fragopanagos, Kockelkoren, & Taylor (2005)
Locus coeruleus-norepinephrine model	Nieuwenhuis, Gilzenrat, Holmes, and Cohen (2005)
The global workspace model	Dehaene, Sergent, & Changeux (2003)
Boost and bounce theory	Olivers & Meeter (2008)
Episodic simultaneous type/serial token model	Bowman & Wyble (2007)
The attentional cascade model	Shih (2008)
The threaded cognition model	Taatgen, Juvina, Schipper, Borst, & Martens (2009)

3.3.3 Locus coeruleus-noradrenaline system

The locus coeruleus (LC) is a nucleus located in the brainstem and the main hub of noradrenaline (NA) synthesis in the brain. The LC is known to be active when an organism is engaged mentally as well as during (relatively) simpler functions such as sleep-wake cycle and arousal (Sara, 2009; Aston-Jones & Cohen, 2005). There are two modes of LC activity, a phasic mode (exploitation) that is evoked by task-relevant or salient stimuli while filtering irrelevant information, and a tonic mode (exploration) associated with disengagement from the task/stimuli (Aston-Jones & Cohen, 2005). The trade-off between the exploitation and exploration states are suggested to optimize performance in behavioural tasks.

LC-NA system and the AB

Given that the LC phasic response acts as a temporal attentional filter for important or infrequent stimuli (Aston-Jones & Cohen, 2005; Dayan & Yu, 2005), LC-NA system was suggested to be involved in modulating temporal selective attention and hence noradrenergic modulation may be involved in the AB effect (Nieuwenhuis et al., 2005, 2007). More specifically, the neurocomputational model of AB (Nieuwenhuis et al., 2005) holds that the AB results

from a temporary unavailability of the LC-NA system following each phasic LC response. The timing of LC refractory period, peaking at around 50 – 100 ms after phasic LC response and lasting for about 400 – 450 ms, coincides remarkably well with the timing of the AB phenomenon (Nieuwenhuis et al., 2005). However, two pharmacological studies directly testing the modulation of AB in human subjects with noradrenergic drugs that reduces NA release have provided mixed results. Nieuwenhuis et al. (2007) reported non-significant results for the difference between the drug clonidine, α 2-adrenergic receptor agonists, and placebo on a visual AB task with neutral stimuli (i.e., letters and digits). According to the authors, a likely candidate to explain these negative findings was that the experimental design and task may not sensitive enough to reveal the differences⁵. In another study, De Martino et al. (2008) observed attenuated AB by administering 40 mg propranolol, a β -adrenergic antagonist, regardless of the valence of T2 stimuli (i.e., emotionally arousing and neutral words). Critically, they also found that the emotional T2s, which are supposed to elicit an increased release of NA, were detected better than the neutral T2s even with the propranolol administration. Based on these findings, De Martino et al. (2008) proposed that T2 detection (where T1 is reported correctly) is proportional to the magnitude of NA release, which was assumed to be greater for emotionally salient T2s.

LC-NA system, mental effort, and the pupil

In terms of methodological relevance to Paper III, the pupil is considered as a physiological index of LC activity, thus studying pupillary responses during attentional selection and allocation of cognitive resources can indirectly inform us about the LC-NA system's activity (see Laeng et al. 2012; Laeng & Alnaes 2019 for further readings on the relationship between LC-NA system, mental effort and the pupillometry method).

The activity of LC, along with that of the hypothalamus, is crucial for regulating pupil dilation, as the subcortical pathway leading to the muscle that dilates the iris starts in these structures (Aston-Jones & Cohen, 2005; Samuels & Szabadi, 2008; Szabadi, 2012; Joshi et al., 2016; Mathôt, 2018). Several studies have suggested an association between LC activity and pupil dilations. The most convincing evidence for this link comes from the studies in monkeys and in rodents combining direct recordings of LC-neurons with pupil measurement (e.g., Aston-Jones & Cohen 2005; Joshi et al. 2016; Rajkowski et al. 1993; Reimer et al. 2014, 2016, but see also Megemont et al. 2022 for a contrary perspective). Figure 3.2 illustrates the similarity between LC activity and the changes in pupil size (Aston-Jones & Cohen, 2005).

⁵In the original paper, several other explanations were also discussed. These included the timing of the drug administration (hence the effect of clonidine peaking at a later time in cognitive task), the administered dose of the drug, as well as the possibility that the AB may be not mediated by the LC-NA system. The authors have raised arguments against the possibility of these explanations.

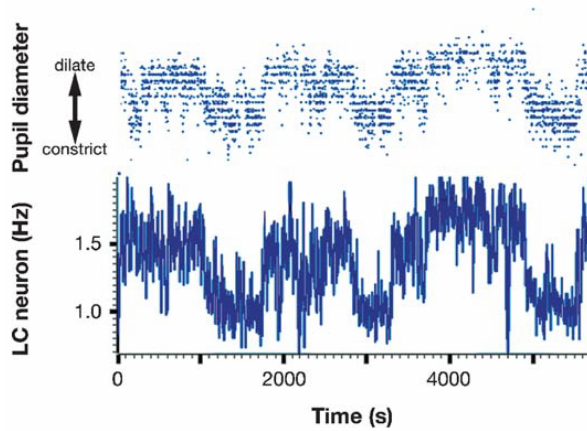


Figure 3.2: Simultaneous recording of the LC-neuron and pupil diameter in monkey. (Figure reprinted with permission from Aston-Jones & Cohen (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, 28, 403–450.)

In cognitive psychology, pupillary changes are also known to reflect mental effort, attentional resources, or the intensity of attention (Kahneman, 1973; Beatty, 1982). For example, a vast number of studies have shown increases in pupillary dilations in relation to increased cognitive demands of a task (e.g., Hess & Polt 1964; Kahneman & Beatty 1966, 1967; Laeng et al. 2011). In support of this notion, in humans, using non-invasive methods (e.g., fMRI), a number of studies have shown that pupillary changes to varying amount of mental effort can indirectly index the LC-NA activity during visual cognitive tasks (e.g., Alnæs et al. 2014) and also in music listening and performance context (Endestad et al., 2020). In addition, usefulness of the pupillometry methodology has been shown in visual studies with a single target detection within a RSVP stream (e.g., Privitera et al. 2010) and also in relation to visual AB tasks (e.g., Wierda et al. 2012; Zylberberg et al. 2012; Willems et al. 2015b).

3.4 Modulating the attentional blink

The year 2005 marks the beginning of a paradigm shift for the theory and discoveries within AB research (Martens & Wyble, 2010). This is because counter-intuitive findings were observed in several studies after this point, particularly in the visual domain, calling the earlier notion of understanding AB as a fundamental, inescapable limitation of human cognition into question. Examples of these counter-intuitive findings include attenuation of the AB effect as a result of task- and stimulus-related manipulations, as well as individual and group-level differences in AB. In the following subsections, I will briefly explain

how researchers have been able to modulate the AB in these ways, covering roughly what we know from the visual and auditory modalities.

3.4.1 Tasks we ‘blink’ less to

Several studies, though mostly in the visual domain, have shown a reduced AB through various task-related manipulations including, but not limited to, simultaneously engaging in task-irrelevant mental activities (e.g., listening to music or thinking about their holiday, Olivers & Nieuwenhuis 2005), performing a secondary task (e.g., responding to a moving dot, Taatgen et al. 2009; detecting a red dot, Wierda et al. 2010; remembering random line patterns presented prior to the presentation stream, Olivers & Nieuwenhuis 2006), and providing different task instructions (e.g., instructions to forget T1; Taylor 2018, instructions to focus attention to a specific time interval in the stream; Shen & Alain 2011). The idea of eliminating or attenuating AB through training participants in the task have been supported in some (e.g., Choi et al. 2012; Tang et al. 2014; Willems et al. 2015a) but not in other studies (e.g., Braun 1998; Taatgen et al. 2009).

3.4.2 Stimuli we ‘blink’ less to

Interest in overcoming the AB led also to empirical studies that involved stimulus-related manipulations. Among these, the experiments on the nature of T2 have taught us important lessons. Shapiro et al. (1997a) was first to report that certain high-priority stimuli can escape the AB. In this study, when a participant’s own name, as compared to other names, was presented as T2 stimuli the AB was no longer observed. Challenging this finding, others have shown that personal names do not always escape the AB (Giesbrecht et al., 2009). Based on load theory (Lavie & Tsal, 1994; Lavie, 2005), Giesbrecht et al. (2009) proposed a ‘load hypothesis’, that is, the perceptual load of T1 modulates the post-perceptual processing during AB. To test this hypothesis, they manipulated T1-load in various ways and found that visual AB was present for personal names in high-, but not in low-load conditions. These findings suggested that increasing T1 perceptual load can place limits on the processing of personal names during AB.

Of particular interest are the AB findings related to face stimuli. Viewed from an evolutionary angle, one would expect faces to be less likely show an AB effect, relative to stimuli with less evolutionary significance.

In nine experiments, Awh et al. (2004) showed that discrimination of face T2s can be performed without any temporal impairment in an AB task. Comparing face and letter T2s in two separate experiments, first, they established that temporal processing of face and letter T2s (when T1 was a digit) differed. AB was observed for letters, whilst face T2s were reported equally well under conditions where T1 were to be reported or ignored. This finding remained also when using a within-subjects design where T2 was randomly varied and exposure duration was adjusted. Even when it was perceptually more difficult to detect faces than letters, T2 deficit was not observed for human face stimuli. Reversing the order of the stimuli (i.e., face T1, digit T2), however, was shown to severely disrupt

3. Temporal limits of attention

the conditional accuracy of T2, gradually recovering as the SOA increased up to around 700 ms. According to Awh et al. (2004), this was due to the presence of a separate processing channel used when discriminating faces, as opposed to competing for resources within a single channel. This is called the multi-channel hypothesis. An alternative explanation for this is the salience of faces, which potentially causes faces to capture attentional resources more effectively than other stimuli. This was tested using a separate set of faces as masking stimuli for T2. Since the processing of face stimuli remained unimpaired, they ruled out the stimulus-driven explanations for the earlier AB results. In another follow-up experiment where both targets were faces, AB was observed. In line with the multi-channel hypothesis, the authors suggested that this is because face T1 occupied the configural (i.e., holistic) processing channels that would be used for face discrimination in T2 task.

Exploring the role of familiarity in identifying face T2s, Jackson & Raymond (2006) conducted three experiments using a visual AB task in which abstract pattern T1s and familiar or unfamiliar face T2s were presented among distracters, which also were either familiar or unfamiliar faces (i.e., making the appearance of face less unique). In this study, independent of the distracters, AB was observed for unfamiliar face T2s, but not for famous face T2s. Opposing Awh et al.'s notion of special configural channels for faces, they suggested that unfamiliar face identification does require attentional resources, similarly as for non-face stimuli. On the other hand, they argued that highly familiar stimuli need less attention and benefit from solid and resilient WM representations. Thus, familiar faces may be immune to AB effects (Jackson & Raymond, 2006).

Landau & Bentin (2008) extended Awh et al.'s finding by showing that faces can escape the AB when compared with objects (i.e., watches), when T1 belong to another category (i.e., flowers). An AB effect was observed when both targets were faces, but its magnitude was much smaller than that of when both targets were objects. In addition, making changes to the T1 task with face stimuli led to an increase in AB magnitude for faces. Here, a face-race discrimination task was selected as T1 task to increase the attentional demands. They concluded that the observations of the lack of AB for face T2s was not due to a perceptual channel for faces, but instead reflects the perceptual salience of face stimuli, which may lower the resources required for detecting faces among other category distracters.

Though I will not go into the details of these studies, the modulation of temporal attention through emotional faces has also been investigated in several studies using the AB task (e.g., de Jong et al. 2009; Miyazawa & Iwasaki 2010; Maratos 2011; Pecchinenda et al. 2020, and using what is known as the emotional AB task (for a review see McHugo et al. 2013).

Taken all together, by demonstrating that certain stimuli *can escape* the AB, these studies provide some evidence in support of the post-perceptual processing of T2 during AB. At the same time, they also indicate that these stimuli do not *always* escape the AB, as the context under which these stimuli are presented seems to matter. This makes it all the more interesting to study how the processing of voices, as 'auditory faces', would behave within a dual-target AB

task as well as in the single-target detection task.

3.4.3 Individuals who do not ‘blink’ or ‘blink’ less

An intriguing set of findings has emerged in the literature suggesting that some individuals are able to override the AB effect. Researchers have distinguished so-called “non-blinkers” (i.e., a group of participants who do not show an AB) from “blinkers” (i.e., individuals who show an AB) (e.g., Martens et al. 2006; Martens & Valchev 2009). These non-blinkers seem to be able to successfully report both targets, independent of the inter-target lag.

Feinstein et al. (2004) was first to report a group of individuals who were able to escape the visual AB effect. In their fMRI study, non-blinkers showed more activation in the anterior cingulate, medial prefrontal cortex and frontopolar cortex compared to blinkers. Martens and colleagues (2006) further explored the question whether fronto-parieto-temporal attention network functions are more efficient in some individuals than in others through measuring brain activity with EEG. Their analysis on the ERP component associated with WM updating showed that the consolidation of targets, in particular, of T2, was slower in blinkers than in non-blinkers. When T2 appeared at Lag 3 (relative to when it appeared at Lag 8), consolidation for both targets was slowed for both groups.

In another visual AB study, Dux & Marois (2008) reported that an individual’s ability to inhibit distracters (as measured indirectly through T2 priming) can predict the magnitude of their AB. With a more direct investigation, Martens & Valchev (2009) showed that non-blinkers are better able to ignore distracters (i.e., digits) than blinkers when identifying targets (i.e., letters). More precisely, when distracters were present in the RSVP stream, the AB magnitude of the non-blinkers was larger (but still rather small) compared with the condition with no distracters. However, the findings with the blinkers showed a substantially larger AB magnitude in the presence of the distracters compared to the absence of them.

More recently, using pupil dilation deconvolution method, (Willems et al., 2015b) showed that the size of the (visual) AB effect can predict the timing of attentional allocation to T2. In this study, participants with smaller AB size were found to allocate attention to T2s faster than the participants with larger AB size. To summarise, the studies on non-blinkers suggest that these individuals update information into WM faster, attend less to distracters, quicker at allocating attention to T2s, and are better able to inhibit target-irrelevant information (Willems & Martens, 2016; Willems et al., 2015b). What remains to be resolved from these studies is who these non-blinkers actually are (e.g., what, if any, are their common traits, experiences etc.?).

One proposal could be that the lack of AB in non-blinkers represents differences in WM capacity. In the literature, the association between WM and AB magnitude has not been found consistently (Martens & Wyble, 2010). Some have observed a negative relationship between AB size and WM operation span (Colzato et al., 2007; Arnell et al., 2008), but not with the backward and forward digit span tasks (Arnell et al., 2008). Others have reported no statistically

3. Temporal limits of attention

significant correlation between WM capacity and the size of the AB (Martens & Johnson, 2008). In another study, it was found that although visual WM capacity did not predict AB magnitude, poor filtering efficiency (i.e., the degree of admitting irrelevant information into visual WM) correlated positively with AB magnitude (Arnell & Stubitz, 2010). Further, AB magnitude may be associated with personality (Troche & Rammsayer, 2013). Impulsivity, a personality trait that refers to the tendency to behave with little forethought (Dickman, 1990), was suggested as a likely candidate to be related to AB. Based on the earlier reports of more efficient information processing among non-blinkers (e.g., Martens et al. 2006) and the association between high impulsivity and difficulty in sustaining attention and less efficient inhibition of irrelevant information (e.g., Dickman 2000), Troche & Rammsayer (2013) tested the association between impulsivity and the magnitude of AB, and whether this trait may differentiate non-blinkers from the blinkers. In this study, non-blinkers were found to be more functionally impulsive (as opposed to dysfunctional) than blinkers (Troche & Rammsayer, 2013). Functional impulsivity is described as a component of impulsivity with positive consequences, rather than being a source of difficulty for the individual (Dickman, 1990). In a preceding study with a non-clinical adolescent sample, it was shown that individuals with higher impulsivity scores had greater AB magnitude (Li et al., 2005). In contrast to Li et al. (2005), Troche & Rammsayer (2013) used a scale that distinguishes between two components of impulsivity (functional and dysfunctional) and compared non-blinkers with blinkers. Although it is difficult to evaluate the role of each component in these two studies, impulsivity seems to be associated with the AB in the visual modality. Impulsivity remains a potential predictor that deserves further attention, also in the auditory modality (as the AB task requires inhibition and selection processes regardless of modality).

In addition to impulsivity, dispositional affect (MacLean et al., 2010) and the personality traits extraversion, openness, and neuroticism (MacLean & Arnell, 2010) have also been shown to predict individual differences in visual AB magnitude. More precisely, MacLean et al. (2010) showed that participants' positive affect was negatively associated with AB magnitude, while negative affect was positively associated with AB magnitude. MacLean & Arnell (2010) found that while extraversion and openness to experience were negatively linked with AB magnitude, neuroticism was related positively to AB magnitude.

Another way to study participant-related variations in the AB effect is through comparing groups of individuals. For example, several researchers have reported attenuation of the visual AB effect in certain groups, in relation to their extensive experiences and/or training in a particular domain. These include, for example, investigations on long-term video-game players (e.g., Green & Bavelier 2003; Wong & Chang 2018), meditation practitioners after intensive training or long-term experience (e.g., Slagter et al. 2007; van Leeuwen et al. 2009), and also experts in relation to their objects of expertise (e.g., car experts and AB for cars, Blacker & Curby 2016).

Finally, an important distinction is that the non-blinkers in the visual modality seem to not show necessarily the same temporal limitations in the auditory

modality (and vice versa). This was shown by Martens et al. (2009) with the observation of a clear auditory AB effect in the visual non-blinkers. Later, it was also found that individuals (in this case musicians) with a large visual AB effect showed a reduced auditory AB (Martens et al., 2015). Returning to the modality debates from earlier, together these two studies indicate modality-specific, rather than central, constraints for the AB phenomenon.

3.4.4 Modulating the auditory attentional blink

The idea of modulating the AB is not completely new in the auditory modality. Previously, modulations of the auditory AB were explored through task-related manipulations. For example, using simple tones, Shen & Alain (2011) investigated whether task instructions for temporal attention-orienting can modulate auditory AB and its neuroelectric correlates as indexed by ERPs. In contrast to the standard AB task, participants were instructed to focus their attention to a specific time interval within the RSAP stream. In the T2 task, they were asked to report whether T2 occurred during this specified time interval, and disregard similar stimuli at other temporal positions. An auditory AB was present when T2 was presented at Lag 1. Percentage of correct reporting of T2 increased when it occurred at the Lag that aligned with the instructions (i.e., attend to short, middle or long intervals). P3b latency, which is a sub-component of P3 (or P300) ERP component, also was shorter when T2 was attended than when it was ignored, with the exception of Lag 1. According to the authors, this shortening of the P3b suggests that short-term consolidation of T2 was facilitated by the deployment of attention to specific temporal positions.

In another ERP study, Shen & Alain (2012) explored whether implicit attentional orienting can attenuate the auditory AB effect within a stream of simple tones. The idea was that by making the T2 more or less likely (80% or 20% probability, respectively) likely to occur in certain positions, people would form temporal expectations of T2. Thus, more attentional resources would be allocated to higher probability positions, which then would improve the detection of T2. Their results showed a significant auditory AB both when the probability was 20% and 80%, but it was reduced when T2 probability was 80% at +2 position (i.e., Lag 2). There was also an increased P3b activity, which may reflect increased attentional resources when participants implicitly orient their attention towards the T2 position.

There have also been a few investigations examining the effects of participant-related factors on the auditory AB. For example, comparing young (mean = 21.2 years old) with older adults (mean = 66.8 years old), with normal hearing, Slawinski & Goddard (2001) investigated age-related changes in the auditory AB task using tonal stimuli (i.e., pure tones). Their findings indicated an auditory AB effect in both groups. Significant group differences in detecting T2 under the dual-task condition were observed at short inter-target lags (Lag 1 to 5, corresponding to SOAs = 90 to 450 ms) but not in the longer lags (Lags 6 and 7, SOAs = 540 and 630 ms). Both in the short lags, and overall, T2 was better detected by young than by older adults. The magnitude of the auditory AB effect

3. Temporal limits of attention

was also significantly larger among older adults in the short lags. Consistent with the inhibition account of Raymond et al. (1992), the authors interpreted these differences between the age groups to reflect less efficient inhibitory functioning among older participants.

Using the same stimuli, another study investigated whether there might be compensatory abilities in auditory attention among congenitally blind compared to sighted individuals (Goddard et al., 1998). Their stimuli included rapid streams of pure tones, among which louder pure tone targets with either low, medium, or high pitch were placed. The task in the experimental condition was to first classify T1 terms of pitch (low, medium, high), then perform the same task for T2 if the tone was heard. In the control condition, the participants performed only the T2 task. The comparison of these conditions across the two groups revealed an auditory AB effect for both groups, but the AB was attenuated among the congenitally blind when the T1-T2 interval was shortest (i.e., Lag 1, SOA = 90 ms). This result was interpreted to indicate some support for an attentional compensation among congenitally-blind.

Another critical stimulus-related manipulation that can potentially modulate the AB is the duration that each stimulus is presented for. Varying durations (ranging from 30 to 160 ms) are used in the existing AB literature, presumably in an attempt to make the task difficulty comparable to that of the visual modality or to make the stimuli intelligible. As shown by Arnell & Jolicoeur (1999), failure to observe an AB in some of the auditory and the cross-modal AB tasks could be explained by relatively long stimulus presentation durations.

All in all, these studies illustrate that the magnitude of the auditory AB effect can be influenced by several factors, including certain task-related factors (i.e., explicit and implicit task instructions), participant-related factors (i.e., age, blindness), and stimulus manipulations (i.e., duration). The comparison of different types of auditory stimuli, their potential interaction effects with duration manipulations, as well as explorations of other participant-related factors can be useful to understand this phenomenon further.

3.5 Musicality and musical expertise

In this final section, I turn to musicality and musical expertise and present how they may modulate the limits of auditory attention. As mentioned in Chapter 1, defining musicality and musical expertise is inherently difficult partly because, given the variety of views on these constructs, no single definition can be agreed upon. From a biological view, musicality can be understood as an innate capacity for music rooted in and constrained by human cognition and its underlying biological mechanisms (Honing, 2018). Alternatively, musicality can be viewed as a capacity for culture unique to humans, determined not through genetics but shared social and cultural ways to understand the world and one another (Cross, 2008).

In disciplines such as music psychology, auditory cognitive neuroscience, and systematic musicology, researchers aiming to understand the potential benefits

of musicality (and musical expertise) has focused largely on comparing musicians and non-musicians, usually by defining musicianship based on years of training, intensity of practice and/or professional involvement in the domain of music. Musicians have been most extensively studied as auditory experts (Chartrand et al., 2008).

3.5.1 Benefits on music-related abilities

Several studies have shown evidence for enhanced sound processing (for a review see Tervaniemi 2009) as well as auditory-motor interactions for musicians (see Zatorre et al. 2007 for a review). I will collectively refer to these enhanced abilities emerging in the auditory domain (excluding speech perception) as music-related benefits for musicians. Empirical data is available on the music-related benefits, covering not only the production and perception of music but also the perceptual dimensions of timbre, pitch, and rhythm. In the next section, I will briefly cover some of the evidence on timbre perception since this is the most relevant for the investigations in this thesis.

A number of studies have demonstrated that musicians' timbre perception abilities are superior to those of the non-musicians. For instance, in a behavioural task where same or different judgements on timbre and pitch dimensions were to be made, Pitt (1994) found that musicians outperformed non-musicians when timbre varied (i.e., trumpet or piano) independent of whether pitch varied or remained constant. In a seminal paper by Pantev et al. (1998) an enhancement in the auditory evoked brain responses to piano tones, as compared to pure sine tones, was observed for conservatory musicians compared with non-musicians with no training. In a follow up study, enhanced auditory cortical representations were observed for instrumental timbres (i.e., trumpet and violin) compared to pure sine tones. These enhancements were preferentially associated with timbres that belong to the musicians' own instruments (Pantev et al., 2001). This provides evidence in support of timbre specificity, that is, "use-dependent [neural] plasticity when musical training has been given on one instrument but not another" (Pantev et al., 2001, p. 172).

Most relevant to this thesis, Chartrand & Belin (2006) examined the extent to which such a benefit of musicianship, as observed in musical timbre perception, can be generalized to timbre from other sound types. In this study, participants were to judge, as quickly and as accurately as possible, whether the sound that they heard belonged to the same or different instrument or the same or different person's voice. Their results indicated that musicians (i.e., minimum 3 years of regular practice on a musical instrument or singing) were better at both the voice and the instrument discrimination tasks, but had a greater advantage in the instrument discrimination task. In terms of RTs, musicians took a significantly longer time to respond than the non-musicians and this difference was more prominent in the voice discrimination task. The authors interpreted the accuracy results to suggest that musical expertise with instrument timbres either generalizes to human voice discrimination, or that musicians have better mental abilities. They explained the slower RTs of the musicians with the possibility

3. Temporal limits of attention

of sound processing at a deeper level, or a difference in the cognitive strategies used by the musicians and non-musicians.

3.5.2 Benefits for cognitive functions

There has been a widespread interest in the possible benefits of musical expertise and musicality for cognitive functions⁶ unrelated to music-related abilities (i.e., far-transfer of the skill). Central to the topics explored in this thesis, a recent review points to a relationship between musical experience and enhanced cognitive functions, including but not limited to, executive functions, verbal memory, attention, and processing speed (Benz et al., 2016). The literature suggests that music training can have long-term benefits for cognitive functions, but the mechanisms behind these benefits remain unanswered (Schellenberg, 2005). General intelligence and executive functions have been suggested as likely candidates to explain the wide range of associations between music and cognitive abilities reported in the literature (Schellenberg & Peretz, 2007).

In a study where differences between musicians' and non-musicians' mental abilities were explored using a large set of tasks measuring different aspects of intelligence, musicians (i.e., participants with an academic degree in music and at least 14 years instrument training) showed a superior performance on the auditory verbal memory task (Brandler & Rammsayer, 2003).

With regards to musicianship benefits for WM capacity, studies have shown that adults and children who received musical training outperform control groups without any training in music on WM measures such as digit span (e.g., Lee et al. 2007; Fujioka et al. 2006). Examining the neural aspects of WM, George & Coch (2011) compared musicians (i.e., participants who studied music for at least 9 years, started playing an instrument prior to 10 years of age, continued practicing the same instrument, and reported actively studying music) and non-musicians (i.e., either did not study music or studied for less than 5 years prior to 14 years of age). Behaviourally, music training has been found to be moderately associated with improvements in WM. Electrophysiologically, quicker WM updating, as indexed by a shorter latency of the P300 component, was found among musicians in both visual and auditory modalities. In another study that treated musicality as a continuous factor among participants who received varying degrees of formal music training, Slevc et al. (2016) found musicality to be robustly associated with WM updating (but not with inhibition or switching components of executive functions) in both visual and auditory domains. This, together with George & Coch's finding of quicker WM updating among musicians, are interesting findings, since faster WM updating was an aspect that seemed to differentiate non-blinkers from blinkers as discussed earlier.

Musicianship benefits for information processing speed as well as for speech perception in challenging listening environments have also been reported. For example, using standardized neuropsychological measures, Bugos & Mostafa

⁶The studies reported here focused on the positive effects, as this section covers the possible benefits related to musicality. However, this does not mean that there are no reports with negative findings in the literature.

(2011) found an enhanced auditory and visual processing speed in musicians (i.e., participants who received minimum of 6 years of musical training, have musical notation knowledge and a regular practice of minimum 5 hours a week) compared to the non-musician participants. Parbery-Clark et al. (2009) reported that long-term musical training benefited perceiving speech in noise, which was mediated by WM when speech-in-noise was measured with a more difficult test involving long and complex sentences.

3.5.3 Musicians and the limits of attention

Due to mixed evidence, the musicianship benefits for attention remain unresolved (Baumann et al., 2008; Carey et al., 2015). For example, in a study by Strait et al. (2010), an enhanced performance was found in self-categorized musicians (with consistent practice for more than 10 years) as compared to non-musicians (i.e., less than 4 years of music training) in the auditory but not in the visual attention test. Rodrigues et al. (2013) reported a greater ability of professional musicians (i.e., orchestra musicians) in selective ⁷, divided and sustained visual attention tasks as compared with non-musicians. Another study reported only a weak evidence in support of the differences between musicians and non-musicians in attention tasks measuring sustained auditory attention and auditory scene analysis (Carey et al., 2015).

Furthermore, a number of investigations featured musicians in relation to the attentional phenomena described in the beginning of this chapter. For example, using both auditory and visual AB tasks (with spoken and visual letters and digits, respectively), Martens et al. (2015) found that musicians (i.e., participants who had attended lessons in playing one or more musical instruments for at least 4 years and actively played a minimum of 4 hours a week) had benefits only in the auditory modality but not in the visual modality. In particular, the musicians exhibited both a delayed and reduced auditory AB effect. This was proposed to potentially indicate better attentional distribution among musicians in this modality.

Palmer & Drake (1997) examined repetition failures in music, similar to the repetition deafness phenomenon, by contrasting the performances of beginner and intermediate level child (varying from 2 to 8 years of training in music) and advanced level adult pianists (12 to 25 years of training). In this study, advanced level adult pianists showed a lower sensitivity (i.e., showed significantly more errors than chance level) to repeating musical pitches and more repetition failures (i.e., higher percentage of errors to repeating pitches) than the children did. The authors explained these results with differences in understanding conceptual structures, that is, advanced musicians are better able to distinguish important events from those that are structurally less important. Based on this interpretation, in a sense, these failures can be understood as the ‘cost’ that skilled players pay as a result of successfully ignoring less relevant events.

⁷This finding in selective visual attention task was at the borderline of 5% statistical significance threshold, $p = 0.049$.

3. Temporal limits of attention

Studying the inattentional deafness phenomenon in music and how it relates to musical expertise, others have examined whether individuals would notice the presence of an “acoustic gorilla”⁸ (i.e., a guitar solo mixed into Strauss’ *Thus Spoke Zarathustra*) while the primary task was counting timpani beats (Koreimann et al., 2009, 2014). In both studies, musicianship was defined in terms of either having 7 hours of weekly instrumental practice in the course of the last 3+ years or 3 hours of weekly practice in the last 5+ years. These studies showed an inattentional deafness effect under musical conditions, and this effect was present even among the musicians. Nevertheless, musical expertise was found to be a significant predictor of the likelihood of noticing the guitar solo, which was around 5 times more likely among musicians than non-musicians (Koreimann et al., 2014). This musicianship benefit was interpreted to be a result of domain-specific knowledge that allows musicians to more effectively process the auditory scene, and devoting less attentional resources on counting the beats.

Agres & Krumhansl (2008) examined musical change deafness by comparing professional musicians’ and non-musicians’ ability to detect changes in melodies under various conditions (changes in musical structure, tonality, musical interval, metrical position, and note duration). In most conditions, musicians (average training of 43.9 years) outperformed non-musicians (average training of 1.6 years) in detecting changes in melodies, with the exception of the tonality condition. More specifically, when the change happened between a scale tone and another scale tone, it was less often detected (i.e., slightly above chance level) both by non-musicians and professional musicians.

To conclude, although causality cannot be inferred, together these studies generally highlight a possible link between cognition and musical training. Continuing the investigations on whether and to what extent musicality and musical expertise can be beneficial in overcoming attentional limits, both within and outside of the musical context, can help better understand the individual differences observed in these phenomena and the complex relationship between WM and attention.

⁸This refers to what is known as the “invisible gorilla” after the famous psychology experiment by Simons & Chabris (1999), illustrating participants’ failure to notice a person wearing a gorilla costume while focused on a task.

Chapter 4

Methodology

In order to evaluate the evidence for an effect, it is necessary to understand the methodology under which that effect was observed. The purpose of this chapter is to provide an overview and a discussion of the methodological considerations and the choices made in the individual papers. In the following subsections, I describe and motivate the experimental paradigms, various measures, sample characteristics, as well as the statistical approach and the analyses used in the thesis' papers. This chapter also includes a discussion of the considerations related to the experimental design and the statistical approach.

4.1 Experimental approaches

All experiments in this dissertation have utilized the rapid serial auditory presentation (RSAP) paradigm. In Papers I and III, the RSAP paradigm was embedded in an auditory variant of the attentional blink (AB) task. Both experiments that comprise Paper II included the RSAP paradigm alone. Below I provide the rationale for employing these two experimental approaches in the present work and discuss the main differences between them.

4.1.1 Rapid serial auditory presentation paradigm

Analogous to the RSVP paradigm (Potter & Levy, 1969) in the visual modality, the RSAP (also referred to as Rapid Audio Sequential Presentation or RASP, Suied et al. 2013) is a paradigm where streams of rapidly changing brief auditory stimuli are used to study the temporal aspects of auditory processing.

The motivation for using the RSAP paradigm were two-fold. First, the RSAP paradigm enables the study of the temporal processing of, and attending to, brief sounds. Second, utilizing the RSAP paradigm alone in Paper II allowed us to study the effects of duration and timbre in the rapid recognition of a single-target presented amongst other brief sounds, while also helping to separate brief sound recognition from the dual-task nature of the AB paradigm and the aspects related to it (e.g., lag manipulation and task switch) in general. A pragmatic benefit of using this paradigm outside of the AB context was also that some of the design decisions in the experiment in Paper III could be based on the results of the experiments in Paper II. This way, we were able to simplify the experimental design in Paper III by reducing the number of experimental conditions based on these informed decisions.

To the best of my knowledge, only a few studies in the literature (Isnard et al., 2019; Suied et al., 2013) have studied the processing and recognition of timbre in brief sounds using this paradigm. While studying brief sounds in isolation

can still be a valid alternative depending on one's aim, this was neither ideal nor desired in the context of this thesis. This is because of the following reasons: 1) Auditory recognition and attention rarely happen in isolation. Anecdotally, we are often required to rapidly process sounds in the presence of other sounds. The decision to utilize the RSAP paradigm (as opposed to studying sounds presented in isolation) both made these studies relatively more similar to how we listen in the everyday context (although, as I noted elsewhere, I make no pretense of claiming the stimuli used in these studies are ecological), while also ensuring that the paradigms in all three papers were similar in terms of the context that the target sounds were presented in, that is, a rapid stream of environmental sounds. 2) Recognizing and attending to sounds presented in isolation is markedly simpler than recognizing and attending to sounds presented amongst other sounds. Accordingly, I expected that the minimum duration required for recognition of sounds presented in isolation would be much shorter (this would also explain the remarkably short duration thresholds for auditory recognition found in the literature) and therefore no longer comparable to our AB studies. 3) I was interested in situations where selection processes are involved (i.e., not only the recognition of a specific sound after the subject was presented one sound at a time, but recognizing a specific sound while optimizing the selection of that sound amongst other sounds). This brings us to the topic of studying auditory selective attention, and to the auditory AB task.

4.1.2 Auditory attentional blink task

The AB task is essentially a two-target RSVP/RSAP procedure originally developed to study the time course of attention (Shapiro et al., 1997b). While the AB is most commonly studied using an AB task embedded in the RSVP paradigm in the visual modality, the existing literature suggests that it is not strictly a visual phenomenon (see Potter et al. 1998 for an opposing view). The auditory AB task involves listening to a stream of auditory stimuli at a rapid pace, with the goal of identifying or detecting two pre-defined targets dispersed among distracter sounds. We constructed two experiments with the auditory AB task (Paper I and Paper III) with the goal of answering the main research questions posed in Subsection 1.2.

The key methodological aspects that differentiate the auditory AB task from the RSAP paradigm are the dual-task requirement and the critical manipulation of lag between T1 and T2. The lag can be defined either in terms of the number of intervening distracter stimuli between the targets or in terms of the time between T1 and T2 onsets, also known as the stimulus onset asynchrony or SOA.

As the readers may have realised by now, given how small the ratio of the AB studies in the auditory modality is relative to that in the visual modality, the methodological framework in our auditory AB studies are, at least to some degree, based on what we know from the findings of the visual studies. According to MacLean and Arnell (2012), the characteristics of a typical (visual) AB paradigm that will reliably produce an AB effect include two targets, a presentation rate of 10 stimuli per second or 100 ms per stimuli (which may vary depending on

the stimuli), requiring an unsped response, including T1 and T2 masks, and containing minimum two lags: one located inside the critical period where we expect a temporary post-T1 impairment (e.g., Lag 2 or SOA < 500 ms), the other outside of the critical period (e.g., Lag 7 or SOA > 500 ms). Nevertheless, this does not mean that what is ‘typical’ of the visual AB studies will necessarily apply just the same in the auditory context. Continuing to experiment is essential to update what we know today as ‘typical’ characteristics of the AB from the visual modality and from the limited auditory AB studies. This is a contribution of the present thesis.

4.2 Behavioural measures

4.2.1 Target accuracy

In those papers using the auditory AB task, we reported two behavioural accuracy measures, namely T1 accuracy and T2|T1 accuracy. T1 accuracy refers to the correct identification of the first target. T2|T1 accuracy refers to a conditional accuracy of T2, given that T1 is correctly reported. Hence, measuring T2|T1 accuracy requires one to first isolate the cases where T1 was correctly identified. These data were used to calculate the rate of accuracy by dividing this by the number of trials under each condition (i.e., max. count of correct response). Accuracy measures alone are not enough to determine the AB effects. The presence of an AB effect is defined in terms of a reliable impairment of T2|T1 accuracy in the short lag compared to the long lag. Thus, statistical tests were run to determine the effect of lag (as well as the interactions of lag with the other effects of interest) on T2|T1 accuracy.

4.2.2 Target recognition sensitivity

In the experiments in Paper II, a sensitivity index (d' or d-prime) based on Signal Detection Theory (Green & Swets, 1966) (or as more often referred to nowadays, as Detection Theory; Macmillian & Creelman 2005) was used as a behavioural measure of target recognition. Since the task in these experiments tested recognition of stimuli based on two types of trials (where a target sound was either present or absent in the RSAP stream) and binary response alternatives (target present and target absent), using a sensitivity measure was preferred. This measure indicates the strength of the target recognition ability, by taking into account how well the participant can maximize the rate of hits (by having more hits and/or fewer misses) while minimizing the rate of false alarms (by having more correct rejections and/or fewer false alarms) in a given condition.

The equation for calculating d' (Macmillian & Creelman, 2005) is shown in Equation 4.1. To this aim, first I calculated Hit rates (hits/(hits + misses)) and False alarm rates (false alarms/(false alarms + correct rejections)) under each condition for each participant. Next, I applied Z-transformation for these two rates in all conditions and participants using the NORMSINV function in Excel.

4. Methodology

Finally, I subtracted the Z -transformed false alarm rates from that of the hit rates to obtain the d' parameter.

$$d' = z(\textit{Hit rate}) - z(\textit{False alarm rate}) \quad (4.1)$$

4.2.3 Maximal AB and AB magnitude

The measures relating to the size of an AB effect are most commonly used among studies with an emphasis on individual differences in the AB phenomenon. Using these measures, it is possible to obtain a single AB size score for each individual. In this work, to measure the size of an AB effect from each participant, two measures are used: maximal AB (Paper I) and AB magnitude (Paper III). I chose to refer to them with different names, since they are similar but not identical. These measures were used because, in addition to the AB effect and its modulations, the research questions in this thesis were also concerned with how certain participant factors may relate to the AB phenomenon.

Taking individual differences into account, the size of the deficit in detecting T2 within the critical AB window can vary among individuals, that is, for one individual, the size of an AB can technically be larger if we compare Lag 2 with Lag 9 (than say, Lag 3 and Lag 9), but for another, it can be still larger when the AB is measured from Lag 3 to Lag 9. The maximal AB measures the size of an AB effect where it could be largest for an individual and was calculated with the following equation in Paper I, similarly to what was suggested in a previous study by (Colzato et al., 2007): $T2|T1_{\text{Lag9}} - T2|T1_{\text{min}}$, where I subtracted the minimum T2|T1 score in one of the short lags (i.e., whichever short lag yields the lowest T2|T1 for a given individual) from the T2|T1 accuracy at long lag (i.e., Lag 9). As we tested multiple short lags in this study, we had the opportunity to adjust the calculations based on where the AB was largest for each individual. This was one of the advantages with using this measure. In Paper III, we only manipulated two levels of lag (one short, one long lag), therefore using the previous equation was not an option. Instead, I calculated the size of AB, referred to as AB magnitude, using the following equation: $T2|T1_{\text{Lag9}} - T2|T1_{\text{Lag3}}$.

I would like to point out that for both measures we can only talk about the size of an AB effect within the limits of the tested lags. That is, it is plausible that an individual would show a larger AB had we used instead, or also manipulated, say Lag 7 and Lag 2. This is a limitation, especially for Paper III. Including a full range of lags would also make it possible to provide the dynamics of time course of auditory selective attention. While testing more lags may be ideal for the AB studies with simpler designs, in the present studies it was not feasible since this would have increased the number of conditions substantially.

4.2.4 Response time

Response time (RT; also called as reaction time) is a measure that has been used in psychology since the late 19th century. The idea behind the RT measure

is straightforward: the more difficult the information processing is, the longer RTs are. RTs can also be useful to provide information about natural stimulus processing when recognition accuracy is very high Agus et al. (2012). Complex RTs (i.e., in response to multiple stimuli, based on two alternative choices) were used in the experiments in Paper II as a dependent variable. In addition, in this study, we wished to examine whether RTs would co-vary with the other measures of interest.

As discussed in Paper II, some of the choices made in relation to this measure may have confounded the RT results. First, as a result of using keyboard as a response device, some delay and variability was introduced into the RT data. A response box would have offered the most precision. Second, in both studies, the participants were instructed to respond as fast as possible while not compromising the accuracy. Speeded responses could have ensured that the RT data reflect the intended purpose, i.e., the speed of response. However, this would have been at the expense of lowering the validity of the recognition sensitivity measure. It would have been perhaps the most ideal to conduct one of these experiments speeded and having the RT as the primary dependent variable, and the other unspeeded and with recognition sensitivity as the main dependent variable. This would have offered better control while helping to separate their underlying processes.

4.3 Pupillary measures

Pupillometry is the study of changes in pupil size that occur as a function of mental activity. In the history of psychology, following the seminal works of Hess & Polt (1960, 1964) and Kahneman & Beatty (1966), pupillometry was established as a reliable and sensitive measure of mental effort. Importantly, this method offers a window into understanding how neural structures related to attention function (Laeng & Alnaes, 2019). In this thesis, pupillary measures were included in Paper III because we were interested in studying mental effort and indirectly the modulations of the LC-NA activity (as discussed earlier in Subsection 3.3.3) during an auditory AB task.

In modern research, pupillary measures are typically obtained using video-based systems primarily designed for tracking eye movements. These eye tracking systems project infrared light towards the eyes. The human eye is blind in the infrared range. The light reflections on the eyes are registered by a camera that is sensitive to infrared light, which are then used to calculate parameters of pupil size from the distance between the camera and the eyes. These modern eye trackers are capable of measuring the pupil non-invasively and with high precision.

Eye tracking systems can be either stationary or head-mounted. Stationary eye trackers are mounted near the screen used to display the experimental stimuli and they record participants' eyes from a fixed distance, while the head-mounted trackers are typically mounted on eyeglasses and record the eyes at a closer range. The former is typically used in laboratory settings, where participants



Figure 4.1: An illustration of the pupillometry experiment setup in Paper III. (Photo: Annica Thomsson)

are required to sit in front of a computer screen, whilst the latter is used more in more natural settings which allow the participants to move around¹.

As a natural consequence of less movement, measurement distortions, such as head movements, are less of a problem in the stationary systems. For this reason, it is argued that the pupil size is best recorded with stationary eye trackers (Holmqvist et al., 2011). In our investigation in Paper III, we used a stationary eye-tracking system (by SensoMotoric Instruments), together with a chin rest (see Figure 4.1 for an illustration). These systems by SMI keep track of head position and distance from screen. In a solely auditory study, on the face of it, it may seem like asking the participants to sit in front of a screen is not necessary. Having a screen-based setup was, however, useful, both for minimizing distortions and noise resulting from variations in lighting (see below), and also for practical reasons such as when giving standard and written instructions (about the study, marking the beginning and the end of different blocks, break points etc.), which helped in minimizing experimenter bias.

Several important principles should to be applied in the studies using pupillometry methodology in order to ensure that the pupil measure reflects the cognitive processes of interest (and not something unrelated). In the study

¹This is more relevant for monitoring gaze, rather than pupil response.

design of Paper III, we paid special attention to these principles.

A major consideration for pupillometry studies is the lighting conditions. Since the pupil is strongly sensitive to changes in light, an important principle of the pupillometry method is that the conditions need to be iso-luminant. Changes in pupil size driven by cognitive factors are typically a lot smaller in magnitude than the luminance related changes (Laeng & Alnaes, 2019). As such, changes in the lighting conditions can introduce significant amount of noise in the pupillary measures. The study in Paper III was run in one of the rooms in our laboratory that is set up for running eye-tracking and pupillometry studies. During pupillometry recordings, it is not preferred to have a very dark room, as this will lead to large and variable pupil size, whereas a bright room generally leads to better quality in pupil data (Holmqvist et al., 2011). Thus, the light conditions in the room we used can be considered as ideal for a pupillometry study. The room is windowless and moderately illuminated from one constant light source at the ceiling.

Furthermore, one should ensure that variations in the pupil response is not due to differences in the luminance of the visual stimuli. This brings us to another related principle, which is having a baseline condition. The baseline pupillary measures, repeated during a task, also provide a monitoring of changes in tonic state of arousal. The study in Paper III was concerned with, as with the rest of the papers in this thesis, auditory modality (thus did not involve any visual manipulation) but we did make use of visuals on screen when providing the task instructions and during the task (i.e., during trials where the auditory streams were presented and when the responses were given). The baseline image involved a simple grey neutral background with the same luminance (i.e., adjusted in their RGB units) as the image presented during the trials and in the response screen. To obtain a baseline measure, we recorded pupil response during the baseline image presentation in silence for 500 ms prior to every trial onset from each participant. We used the “subtraction method”, where the baseline pupil was subtracted from the pupil response during trials, to ensure that the resulting measure (i.e., pupillary change from baseline) reflects the internal mechanisms we intended to measure and was not due to the light conditions or changes in tonic arousal (Laeng & Alnaes, 2019). Combined with the repeated measures (rm) analyses, by using a baseline correction, we were also able to effectively rule out the individual differences due to the effects of various external factors (e.g., medication, coffee) and at the baseline (e.g., anatomical differences) on the pupillary measure.

Calibration errors can also lead to inaccurate measurements of the pupil dilations. To ensure the accuracy of the pupil measure throughout the experiment, we performed 5-point calibration and validation routines prior to each experimental block. According to SMI’s scientific user guide (2017), 5-point calibration offers the best trade-off between the duration of the calibration process and calibration accuracy in most cases.

Finally, pupil size is known to be sensitive to fatigue. This becomes especially important when studying attention. In addition to the trial-related baseline corrections and (partial) counterbalancing of stimuli, in the experiment presented

4. Methodology

in Paper III, various design choices were made to limit the number of trials (e.g., potentially interesting variables/conditions were excluded). In this experiment, the presentation streams in each trial were very short (i.e., 2 seconds), the trials were self-paced (i.e., initiated by the participant, which allowed for adjusting the rate of trials to their own pace), and the pupil recording during the entire study lasted on average around 45 minutes. To ensure the pupil data is not affected from fatigue, we also included 5 break points after completing each block in the experiment. In these breaks, we encouraged participants to take a couple of minutes break if needed and rest their eyes.

For the pupil data analysis, we used mean pupillary change from baseline under each trial and for each participant. More advanced techniques, such as the deconvolution method could have been also useful to explore the time course of pupillary dilations during an AB task, but the mean pupillary change across conditions were enough to run the analyses needed for our study's aim.

4.4 Working memory measures

Since the tasks in all four experiments involved holding material in WM, differences in WM could partially account for task performance. Our interest lay primarily in exploring how WM may relate to the individual and group level differences in maximal AB scores, (Paper I) and to the recognition of brief tonal information (Paper II).

WM was assessed using Letter Number Sequencing (LNS; in Paper I) and Digit Span (DS; in Paper II, Experiment 1b) tests of Wechsler Adult Intelligence Scale (WAIS-III; Wechsler 2003). These two tests were selected for three reasons: 1) They are the most highly correlated psychometric tests to the laboratory tests (i.e., operation span, listening span, *n*-back task) to assess WM function (Shelton et al., 2009). 2) Among all subtests of WAIS-III, they contribute the largest in predicting working memory (Hill et al., 2010). 3) They are both quick to administer.

LNS is a complex WM span task in which participants are presented with an unsorted auditory sequence of number and letter combinations and asked to state the numbers in ascending order and the letters in alphabetic order. The sequences increase in length as one progresses in the test. The total score of an individual is calculated based on the total number of correctly sorted sequences. The scores obtained from the LNS task was used to explore the relationship between WM Span and the maximal AB, both at the group and individual level.

DS is composed of Forward, Backward, and Sequencing subtests, in all of which the participant is presented auditory sequences of digits that increase in length. In DS Forward the participant is asked to repeat them back in the same order, while in DS Backward, the participant is required to repeat the digits in reverse order. In DS Sequencing subtest, the participant is asked to sequentially order the numbers presented. The sum of task points and the longest correctly reported sequences are reported for each subtest. Total DS score is calculated based on the task points obtained in all three subtests.

4.5 Self-report measures

4.5.1 Musical sophistication

In Section 3.5 I covered the ways in which musicality and musical expertise can influence sound processing and cognitive functions such as attention. In all three papers, musicality was evaluated by means of the Goldsmith Musical Sophistication Index (Gold-MSI; Müllensiefen et al. 2014a). The Gold-MSI measures a general musical sophistication score and five domains: active engagement, perceptual abilities, musical training, singing abilities, and emotions. A higher score denotes a higher musical sophistication level. In Papers I and III we used the Gold-MSI in its entirety, while in Paper II we employed the general musical sophistication sub-scale.

Prior to the selection of Gold-MSI, other measures of musicality such as, Ollen Musical Sophistication Index (OMSI; Ollen 2006) and Musical Ear Test (MET; Wallentin et al. 2010) were also considered. While both the OMSI and Gold-MSI are self-report musical sophistication inventories, MET, on the other hand, is a computerized test of musical ability which includes melody and rhythm sub-tests. In comparison with Gold-MSI, OMSI is less comprehensive and centrally focused around music practice and training. In fact, this is the shortcoming that led to the development of Gold-MSI. The MET was considered mainly for Paper I. In retrospect, given the two participant groups (expert versus non-musician) we recruited for Paper I, this paper would have benefited using a test such as MET, which is designed to distinguish professional musicians, amateur musicians, and non-musicians without exhibiting floor and ceiling effects (Wallentin et al., 2010).

As I progressed in my PhD work, my interest shifted more towards capturing aspects of musicality that “do not necessarily arise strictly from the theoretical knowledge or professional involvement in music making” as stated in Paper II (pp. 4) and as a researcher I became more sceptical about conducting research where I would categorise the musical sophistication of individuals based on a cut-off score or a definition. This kind of thinking aligns well with how musical sophistication is conceptualised and implemented in Gold-MSI, as it focuses on the aspects of musical activity that are not limited to musical instrument practice and is developed as a tool to measure musicality among non-musicians (Müllensiefen et al., 2014b). These were the reasons for employing Gold-MSI in Papers I, II, and III. Using the same measurement across all investigations allowed us to have a better control for threats to internal validity due to differences in instrumentation.

4.5.2 Impulsivity

In cognitive neuroscience, selection and inhibition of responses are treated as two sides of the same coin and are mediated by the same brain circuit (Mostofsky & Simmonds, 2008). Since inhibitory processes are related to response selection, as well as top-down control and information processing, having a general tendency

4. Methodology

to make impulsive responses may be related to the performance on a selective attention task with a rapid stimulus presentation.

Due to the theoretical relevance of impulsivity to information processing and to response selection, as well as the previous findings in the (visual) AB literature regarding the relationship between the impulsivity and the AB performance, we decided to include a measure of impulsivity in our final paper. In Paper III, the short version of Dickman's Impulsivity Inventory (DII; Dickman 1990, excluding the filler items) was employed as a measure of the Trait Impulsivity. A higher score represents higher impulsivity. This measure distinguishes functional and dysfunctional components of impulsivity, which was one of the reasons for selecting this measure in particular.

4.5.3 Subjective certainty ratings

Participants' task-related responses in these experiments are only the behavioural outcome. Alone, they do not indicate much about participants' experience of the task. Although it did not go into the published work, I found it valuable to hear about the participants' comments and suggestions after conducting the first AB experiment (Paper I). In the experiments in Paper II, we included a subjective certainty rating measure after each trial in addition to informally ask participants' comments and suggestions as before. Subjective certainty ratings assess the participants' degree of certainty about the responses that they have provided. They rated their response in each trial on a scale ranging from 1 (very uncertain) to 6 (very certain). Such ratings are commonly used in studies applying the Signal Detection Theory, as certainty ratings make it possible to conduct analyses (i.e., receiver operating characteristics) that go beyond binary responses. The main motivation for including this measure in Paper II was both to test if the certainty levels were influenced by our experimental manipulations (e.g., to be able to provide answers to questions, such as, whether the participants were less certain about recognizing a target when the sounds were presented in shorter durations, based on data) and to test if this measure is related to other measures of performance (e.g., *d*-prime and RTs). We also have considered using subjective certainty ratings or a similar (e.g., subjective rating of mental effort) self-report measure in Paper III. In order to not make the experiment longer than it was (since the pupil is sensitive to fatigue), in the end, we decided not to include it.

4.6 Participant recruitment and sample characteristics

All four experiments presented in this thesis are based on data from separate samples. Sampling involved a combination of purposive and convenience sampling methods. Participants were recruited through advertisements (flyers posted on social media and in campus) and personal contact. The details of the recruitment locations and the inclusion/exclusion criteria in each study are described below.

Paper I – Non-musician participants were recruited from the Facebook group “International Students at the University of Oslo (UiO)”. Cellists were mostly

recruited from the Norwegian Academy of Music. Inclusion criteria for both groups were being 18 to 40 years old and having normal hearing. For the non-musician group, we had the additional criterion of having no/very little formal musical training. For the cellist group, the additional criterion was being a professional cello player (e.g., performing artist, teacher, conservatory student). Exclusion criterion was having a history of neurological disorders.

Paper II – Participants in Experiment 1a were recruited from the “International Students at the University of Oslo (UiO)” Facebook group, while in Experiment 1b they were recruited in connection with another ongoing study for the development of a test battery at the cognitive laboratory. These participants joined the present experiment first, and then took the rest of the experiments in the test battery. Inclusion criteria in Experiment 1a were being between 18 to 40 years of age, having normal hearing, and having little or no formal musical training. In Experiment 1b, the inclusion criterion regarding musical training was intentionally made less strict so that we would reach a wider participant profile in terms of musicality. Exclusion criterion in both experiments was having a history of neurological disorders.

Paper III – Participants were recruited mainly from two Facebook groups: “International Students at the University of Oslo (UiO)” and “Kringsjå student village”. Inclusion criteria were being between 18 to 40 years of age, having normal hearing, and having normal to corrected eye vision. Exclusion criterion was having a history of neurological disorders.

To avoid training effects in task (as the tasks and the stimuli were rather similar in all experiments), we aimed not to test the same participant more than once in these experiments. One participant, whose data we excluded from the second study, joined two of our experiments.

4.7 Experimental design considerations

There were several choice points in designing the experiments that comprise this thesis. Some of these concern more general aspects of the experiments:

The first consideration pertains to the experimental design technique. A factorial experimental design was used in all four experiments to study the separate and combined effects of second target type (Papers I, II, III), lag (Papers I and II), stimulus duration (Papers II and III), and participant group (Paper I) on behavioural task performance (Papers I, II, III) or on the amount of mental effort associated with the task (Paper III). The nature of our studies required a factorial design as our research interest lies in the combined effects. Factorial designs are also efficient and cost-effective in indicating the main and interaction effects of multiple independent variables. Another advantage of factorial designs is that they allow for simultaneously studying multiple independent variables, whilst keeping the extraneous variables balanced across the conditions (e.g., as I did by building some of the potential confounds such as target position and first target type into the design). A weakness of conducting a factorial design experiment, as opposed to a one-factor-at-a-time experiment,

4. Methodology

is that as a result of having multiple independent variables in the design, the number of participants required also increases. Another limitation of factorial designs concerns the higher-order interactions. Higher-order interactions (e.g., a three-way interaction effect) can be difficult to interpret. I have dealt with this challenge by carefully examining the post-hoc results, the plots, and sometimes by conducting several split-up analyses.

Within-subjects design was opted for whenever possible, and this was for many reasons; such as, to control for individual differences (as the attentional blink literature seems to support the role of individual differences in this task), to increase sensitivity of the experiments, and to lower the number of participants required, to name a few. Paper I included a factorial design based on a mixed model, with one between-subjects variable and two within-subjects variables. Papers II and III were constructed as factorial designs based on within-subjects variables only. One concern with the within-subjects design, especially when studying attention, is fatigue. We ensured that the length of the studies was not particularly long (no testing lasted longer than 1.5 hours per session, including giving information about the study, filling the questionnaires, and debriefing) and the participants were given opportunity to take short breaks. The main problem with within-subjects design is that testing the same participants multiple times increases the risk of introducing carry-over effects (e.g., learning, habituation, practice effects), which pose a threat to internal validity. This and other sources of bias related to sequencing of the conditions can be effectively dealt with by counterbalancing and randomization.

All four experiments presented in this thesis were built with a block structure and participants went through all experimental blocks (the details of what was manipulated in each block and the number of blocks can be found in the individual papers). Hence, using experimental control techniques on the order of these blocks was crucial for the internal validity of our experiments. In the experiments in Paper I and II, the order of blocks were fully counterbalanced. For this, I wrote an in-line script at the experiment startup in E-Prime, and specified the counterbalancing orders of the blocks by assigning subject numbers. Randomization of the sounds (and their selection from the pool of sounds belonging to distracters and target categories) as well as the locations of targets in the RSAP stream were also programmed in E-Prime in these experiments.

The pupillometry experiment in Paper III was created using the Experiment Center software by SMI. Experiment Center does not allow for adding in-line codes and lacks advanced features for experimental control, but it is possible to randomize the trials. In this experiment trials were randomized using this function. In the preparation of the RSAP streams, we ensured that the order of the distracters were randomized and the targets and distracters selected from the lists randomly. The orders of the blocks in this experiment were partially counterbalanced by creating multiple experiment files (i.e., 4 versions with reversed orders of the block manipulations), but we additionally assessed the effects of serial order effects post-hoc to separate potential carryover effects (should they result from the incomplete counterbalancing) from the experimental effects. The results of this analysis showed that, as we reported in the paper,

the serial order seemed to only have an influence on the accuracy of the T1 identification, where the participants who went through first 30 ms and then 90 ms stimulus presentation conditions did better than the other way around. So, assuming the 30 ms condition is more taxing than the 90 ms condition, in a sense, practicing with the 'heavy sword' first seems to have benefited T1 performance only.

While on the subject of learning and practice effects, when designing these experiments, I also considered the reasons for and against including a practice session at the beginning of an experiment. Inclusion of a practice session gives the participant an opportunity to get familiarised with the task by practicing both what to expect from the 'actual' experiment and what is expected from the participant himself/herself. Having a practice session (especially if one includes an accuracy criterion for completing this session and/or provides response feedback) will surely reduce response errors by ensuring that the participant understood the task, but it also introduces an element of training and learning. The experiment in Paper I included a practice block, while the experiments in Papers II and III did not. In Paper I, we observed performances reaching ceiling level in the experimental sessions (which may or may not be related to the practice; other potential reasons are discussed in the article). The rationale behind not including the practice session in our later studies was to be able to study the recognition and attention to brief sounds without prior practice in the task.

There were also several choice points that required much more specific decisions. These included decisions about the 'quantities', such as the number of total trials in the experiment, the number of items per trial, how many participants to test, how many and which lags to manipulate, how long should a sound stimulus and an inter-stimulus interval last; and about the 'qualities', especially in the preparation of the sound materials, such as the selection procedure of the targets (e.g., which pitches, whose voices, what to contrast the voices with), and of the distracters, truncating procedure (e.g., from which part of the sound signal) and selecting the normalization method (e.g., RMS or peak normalization). I have made these decisions by consulting the literature and my supervisors. While each decision regarding which conditions to include or exclude limited the scope of the investigations, on the plus side, they kept the experiments a manageable size, thereby allowing us to obtain useful data from our participants.

Doing research in times of uncertainty has brought some unique challenges. One such challenge was with regards to the experimental research setting. Although it is common practice to conduct experimental studies in a laboratory setting, due to the rise of a global pandemic, at first, we designed the experiments in Paper II to be conducted in an online platform. While trying to control for the order effects in an online environment, I ran into an engineering problem. After discussing the potential solutions for this issue, we came to realise that this was going to take a long time to solve and also introduce more weaknesses than strengths to the experiments. It was in one of those meetings that Tor Endestad kindly suggested me to instead run this study in the laboratory, together with the help of his research assistants in combination with another experiment going

on in our laboratory. In the experiments presented in Paper II and III, we took extra measures to ensure the participants' health and safety and to follow the rules and regulations by including new safety protocols to these studies.

4.8 Statistical approach and data analysis

4.8.1 Methods of statistical inference

Null hypothesis significance testing (NHST) is a heavily debated yet arguably the most commonly used method of statistical inference in psychological science with a very long tradition (Nickerson, 2000; Krueger, 2001). The long history of debates about NHST, according to Kline (2004), was characterized by a lack of consensus and as such there is no set of recommendations that can please everyone. Nevertheless, as suggested by many (Denis, 2003; Kline, 2004; Harlow et al., 2013), steering our focus from solely on NHST to supplementing it with alternative methods of data analysis (e.g., Bayesian analysis) seem necessary for making better statistical inferences.

My approach in the work presented here was supplementing, as opposed to abandoning NHST model altogether. This is because of the following (mostly practical) reasons: First, I think this serves better the purpose of improving. Second, having trained mostly in NHST tradition, I am still learning about the alternative methods. Third, in my (short) research experience, communicating my research using only one of these models would be limiting (both in terms of the dialog with the broader research community, and in terms of the interpretations of our results). Although, I must mention that before making this decision, I did consider (and in fact tried at the earlier stages of my PhD journey) reporting my results only using Bayesian methods. However, unlike NHST, the use of Bayesian methods has not been particularly common in social sciences, and they are rather new to most practitioners (Etz & Vandekerckhove, 2018).

In all four experiments, traditional, frequentist approach to statistics (i.e., Neyman-Pearson) was combined with the Bayesian approach. This may be rather unconventional, as they are often seen as opposite camps. Nevertheless, I am not alone in positioning myself in the synthesis of frequentist and Bayesian statistical inference methods. For example, I. J. Good reportedly stated:

“I personally am in favour of a Bayes/non-Bayes compromise or synthesis. Partly for the sake of communication with other statisticians who are in the habit of using tail-area probabilities.” (Nickerson, 2000, p.276).

Bayarri and Berger (2004, p.58) also pointed to the usefulness of a combined approach:

“Statisticians should use both Bayesian and frequentist ideas. [...] The situations we discuss are situations in which it is simply extremely useful for Bayesians to use frequentist methodology or frequentists to use Bayesian methodology”

and raised the importance of a methodological synthesis while acknowledging the existing philosophical differences (Bayarri & Berger, 2004, p.78):

“Philosophical unification of the Bayesian and frequentist positions is not likely, nor desirable, since each illuminates a different aspect of statistical inference. We can hope, however, that we will eventually have a general methodological unification, with both Bayesians and frequentists agreeing on a body of standard statistical procedures for general use.”

All in all, I found it useful to combine them to complement one another. In doing so, I was not only able to reduce uncertainty around interpreting non-significant results but I could also disseminate my research to a wider group of scholars. In the next subsection, I will give an overview of the Bayesian approach, introduce some of the key concepts, and discuss the advantages and disadvantages in the context of my work.

4.8.2 Bayesian approach to statistics

An overview

Probability is essential to statistical inference. Named after Reverend Thomas Bayes (1702 – 1761), Bayes’ Theorem provides a formal way to calculate the probability of an event to occur based on prior information related to that event given the data. It also allows us to compute inverse conditional probabilities.

We apply Bayes’ rule in everyday life in many situations. Weather forecast is a typical example. When deciding whether to bring an umbrella before we head out, we calculate the probability of rain by looking out from the window and observing the clouds, by using the information we have on how the local weather has been in this season. We then can use the combined information to make an estimation of the chances that it will rain (or not) given how the clouds look today. We can also later use today’s data to update our initial prediction of the weather to make better decisions in the future.

We can state Bayes’ Theorem mathematically as follows:

$$P(\mathbf{A}|\mathbf{B}) = \frac{P(\mathbf{B}|\mathbf{A})P(\mathbf{A})}{P(\mathbf{B})} \quad (4.2)$$

Where:

- $P(\mathbf{A}|\mathbf{B})$ is the probability of event A occurring, given that B is true
- $P(\mathbf{B}|\mathbf{A})$ is the probability of event B occurring, given that A is true
- $P(\mathbf{A})$ is the probability of event A
- $P(\mathbf{B})$ is the probability of event B

Note that A and B are independent events.

Bayesian key concepts

I will now introduce some of the Bayesian key concepts, as they are relevant in understanding the analyses conducted in this thesis. These concepts are posterior probability, prior probability, Bayesian updating, and Bayes Factors.

Using Bayes' Theorem, we can quantify the evidence for our beliefs about, say, hypothesis θ , given the data we collected. This is known as the *posterior probability* or *posterior* (i.e., $P(\theta|\mathbf{data})$). To compute the posterior, we would use the *likelihood* of obtaining the data given our hypothesis θ (i.e., $P(\mathbf{data}|\theta)$), and the probability of our belief about θ being true, referred to as *prior probability* or *prior* (i.e., $P(\theta)$). We would also use the probability of the data (i.e., $P(\mathbf{data})$) to normalize the posterior. Using the terminology explained thus far, we can then re-state that, according to Bayes' Theorem, the posterior probabilities are proportional to the likelihood \times the prior probabilities. With Bayesian methods, we can also update the prior understanding of the probability of the hypothesis θ , in the light of new information (i.e., data). This process is called *Bayesian Updating*.

Bayes Factor (BF; Jeffreys 1961) is a way of comparing models (e.g., a null and an alternative hypothesis) under the Bayesian framework, and it refers to the likelihood ratio of evidence to arise from the hypotheses. By quantifying the strength of evidence, BF informs us about which model is more likely given the data and varies between 0 and ∞ . In the papers in this thesis, I use BFs for comparison-based hypothesis testing. When interpreting the BF-values, we consulted the conventional guidelines (i.e., labels and cut-offs suggested for evaluating the strength of evidence) adopted from Jeffreys' work (1961), but it is important to note that this does not mean that we are *required* to make any black and white decisions based on BF-values. With BF, unlike with p -values, we do have the possibility of *updating* by adjusting the odds continuously (i.e., as new data emerge). I will discuss the issue of label categories later in the Section 4.8.3

Why go Bayes? Pros and cons of using Bayesian approach to statistical analysis

Some of the advantages of hypothesis testing using the Bayesian approach over the frequentist NHST are summarized below:

1. BF quantifies how much the data supports one hypothesis over another (Dienes, 2008).
2. BF allows for obtaining evidence in support of the null hypothesis (Lee & Wagenmakers, 2014; Dienes, 2014) and can distinguish insensitive data from support for the null (Dienes, 2014).
3. BF allows for observing the evidence as the data accumulate and can eliminate the problem of optional stopping (Edwards et al. 1963; Rouder 2014; Lee & Wagenmakers 2014, but see de Heide & Grünwald 2020).

Of these, the first two points were the main advantages of ‘going Bayes’ while approaching the data obtained from the experiments in this thesis.

The main drawback of this approach is that the BFs are sensitive to the prior distribution (Lee & Wagenmakers, 2014). While acknowledging the importance of the priors, and that using different priors may lead to different conclusions, Bayesians would argue that priors exist also in frequentists statistics, only that they are flat. The difficulty of calculating the Bayes factors is also mentioned as a disadvantage, but nowadays Bayes Factors can be readily calculated using the standard statistical packages of SPSS, JASP, and R.

4.8.3 Statistical considerations

The debates about NHST and the Bayesian approach go beyond the focus of this thesis, therefore it will not be covered here. Instead, I shall discuss the statistical considerations one should keep in mind for drawing accurate conclusions from the findings of this thesis.

NHST can either reject or fail to reject the null hypothesis. One common mistake is to interpret a non-significant p -value as accepting the null. As Altman and Bland (1995) eloquently put it: “Absence of evidence is not evidence of absence.”. As test of significance alone is not enough to make accurate statistical inferences, what is reasonable is to use some of the alternative methods (e.g., Bayesian analysis, effect sizes, power, confidence intervals, meta-analysis) to reduce the uncertainty. In the papers presented in this thesis, we followed up the results obtained from NHST with further Bayesian tests (at least to a large extent). I acknowledge, however, that the secondary analyses (i.e., the correlation analyses in Papers II and III) could have benefited from being followed up with Bayesian correlation tests. In the interest of keeping the papers to a manageable scope, I had to exclude further analyses. However, since these analyses add additional insight to the questions posed in this thesis, here I included these further Bayesian tests (see Appendix A) and will report from these results in the next chapter.

Another relevant common misinterpretation about NHST is the inverse probability error (Cohen, 1994) or Bayesian Id’s wishful thinking error (Gigerenzer, 1993), that is, taking p -value as the likelihood that the null hypothesis is true given the data. This would be an example of an incorrect synthesis of Bayesian interpretation using NHST method. In contrast to p -values, Bayesian approach estimates the posterior probability of hypothesis (both for H1 and H0), not only the conditional probability of the data (Kline, 2004), as shown earlier in this section. In the papers in this thesis, the posterior inclusion probability ($P(\mathbf{inclusion}|\mathbf{data})$) and the change from prior to posterior inclusion probability (referred to as $BF_{Inclusion}$) obtained from the Bayesian analyses (and not the p -values) were used when interpreting the likelihood of the models or hypotheses given the data.

Power is also considered relevant when it comes to statistical inference based on the traditional approach. It is well-known that the validity of the NHST-based statistical inferences increase when statistical power is high. We reported effect

4. Methodology

sizes and power (with the exception of Paper I, in which the lack of sufficient power could explain the inconclusive results). In the Neyman-Pearson approach, power refers to the likelihood that the null hypothesis is rejected when it is false and is influenced by the statistical significance criterion, the effect size, and the sample size (Cohen, 1992). I used one of the most frequently used computer tools (i.e., G*Power; Faul et al. 2007) for power analysis and sample size calculations. Note that this is only a relevant concern for the results obtained with traditional statistics. Some Bayesians would argue that power cannot be calculated based on the data to determine the sensitivity of those same data in distinguishing the null and the alternative hypothesis and is therefore a flawed solution (Dienes, 2014, 2016). I thought conducting a power analysis in the present thesis could nevertheless be beneficial, as it would help mitigate the pitfalls of running underpowered NHST tests and with research planning.

A related topic is the stopping rule for data collection. According to the traditional approach, one specifies the stopping rule *a priori*, typically by conducting a power analysis, while when using the Bayesian approach, one can combine experiments or continue (or stop) the data collection as one wishes in order to firm up the evidence (Dienes, 2008, 2011). As Dienes (2008; 2011) and Wagenmakers (2007) explains, when using p -values it is guaranteed that one will eventually get a significant result if one has no stopping criteria (as a result of inflating Type I Error), as any value between 0 and 1 is equally likely, whereas with BF one can use an infinite number of participants on many experiments and still the BF may not reach 10, as BF is driven towards zero. This insensitivity to the stopping rule is considered a unique advantage of BF over the p -value (Dienes, 2008). Thus, for example, from a Bayesian perspective, I could have continued collecting more data (or stopped data collection earlier) until we have enough evidence in support of H0 or H1. In our experiments, we used the stopping rule, but also provided a sequential analysis to show the development of the data for the main analyses in Papers II and III. In retrospect, Paper I might also have benefited from this analysis.

Moving next to the Bayesian approach and in particular to the critical considerations on Bayes Factors that readers of this thesis should keep in mind. As mentioned earlier, one of the disadvantages of the Bayesian analysis used in this thesis was that the Bayes Factors depend on the prior distribution. In all the papers, JASP's default priors were used and reported in the individual papers. In future studies, using a method based on posterior distribution/Bayesian parameter estimation or conducting sensitivity analysis for the BFs could be useful.

Critique of the BF also includes arguments against the label categories suggested for BF interpretation. For example, Patching (Patching, 2018), illustrates the problem with these categories with a stick figure showing how researchers may jump for joy, feel happiness, despair, annoyance, or surprise depending on the BF category of their results. I see the problem with this, but I think that this problem actually lies in having a biased approach towards doing research (e.g., over-attachment to one's research, believing only certain results should be granted publication), and not in the categories themselves. This 'Dance of

the Bayes Factors' is perhaps based on a similar argument against the p -values, illustrating how the p -values 'dance' and why we should not trust p -values alone to draw conclusions about the data (Cumming, 2014, see Dance of the p -values). While I do agree that the BF should be used and interpreted with care (and ideally supported with further analyses), I also do not consider the problems with p -value and BF as the same.

4.8.4 Data analysis

In this final subsection, I provide a general overview of the statistical analyses conducted in the articles that comprise this thesis. Data processing and statistical analyses, as well as their intended purposes, are explained in each individual article in greater detail.

Prior to the statistical analyses, for each study, data files obtained from all participants during the experimental sessions were merged, data from the WM measures (whenever applicable), and the self-report measures were entered into an Excel sheet, and all data sheets were screened for errors or missing data using descriptive statistics. For each participant, the scores obtained from the behavioural (Papers I, II, III) and pupillary measures (Paper III) under each experimental condition were calculated as explained in the sections above. Preparatory analyses were conducted to assess the assumptions for the statistical tests and to check for extreme outliers. Upon violation of the sphericity assumption, if the epsilon value was smaller than .75, F -statistics were corrected using Greenhouse-Geisser correction, and if it was larger than 0.75 Huynh-Feld correction was applied. Outliers were identified by visual examination of the box-plots and by identifying the cases with Z -scores in excess of 3.29 (Tabachnick et al., 2007), and were mostly dealt with by replacing the score with one unit smaller/larger than the next extreme score (Tabachnick et al., 2007). This method ensured that we keep these data points still as deviants, while effectively reducing their impact. When dealing with RTs, I based the decision on the type of outlier (short versus long RTs), as these are considered to be generated by different processes (Ratcliff, 1993).

All statistical analyses were performed using JASP (JASP Team 2019; jasp-stats.org). In all three papers, the statistical analyses were conducted with the general aim of determining the effects of stimulus factors of interest on task performance and with the secondary aim of testing whether various participant factors of interest were associated with the measures of task performance. To recap, I provide a visual summary of the various factors explored and the measures used in this thesis, shown in Figure 4.2.

Papers I and III aimed at analysing the effects of our experimental manipulations on performance (measured behaviourally in Paper I, and behaviourally and physiologically in Paper III) in an auditory AB task. For this purpose, rm-ANOVAs were conducted on T1 and T2|T1 accuracy scores in both papers. Paper III also included rm-ANOVAs on the additional dependent variable of pupil size (i.e., mean pupillary change from baseline).

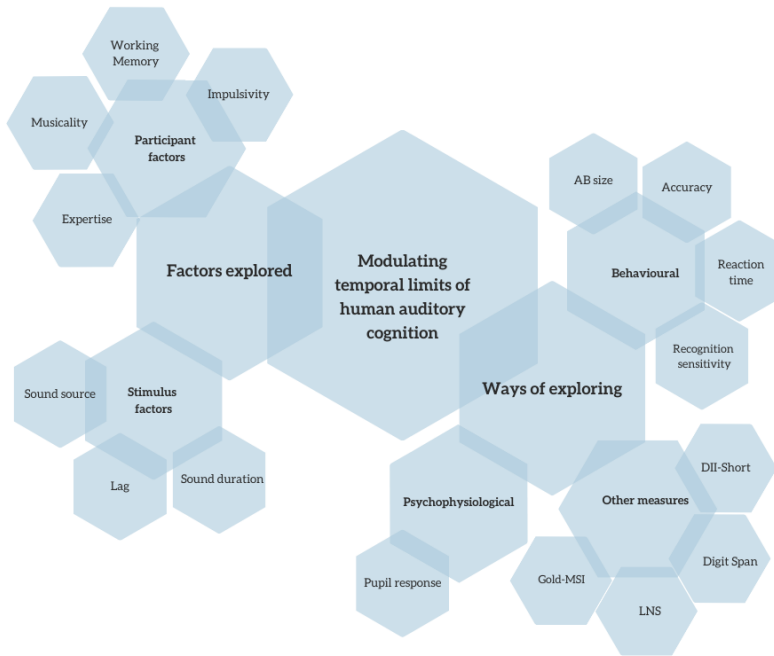


Figure 4.2: A visual summary of the factors explored and the ways of exploration in this thesis.)

In all three papers, further Bayesian rm-ANOVAs were performed to determine which model received most support given our data and/or to reduce uncertainty following non-significant findings. All three papers also included Bayesian paired samples *t*-tests to compare the means between two repeated measures on a dependent variable (e.g., to compare the mean T2|T1 accuracy scores at Lag 3 and at Lag 9). In JASP, it is possible to conduct sequential analysis particularly for Bayesian *t*-tests. Using sequential analysis (in Paper II and III), we showed how the evidence develops across the study sample for the contrasts of interest. Correlation analyses (reported together with the correlation coefficients) were mainly performed to determine if, and how strongly, various participant factors (e.g., musical sophistication, impulsivity scores) were associated with the task-related scores (e.g., AB magnitude, RTs).

Chapter 5

Research summary

“Finding the question is often more important than finding the answer.”

— John W. Tukey

5.1 Paper I

5.1.1 Abstract

Attending to goal-relevant information can leave us metaphorically “blind” or “deaf” to the next relevant information while searching among distracters. This temporal cost lasting for about a half a second on the human selective attention has been long explored using the attentional blink paradigm. Although there is evidence that certain visual stimuli relating to one’s area of expertise can be less susceptible to attentional blink effects, it remains unexplored whether the dynamics of temporal selective attention vary with expertise and object types in the auditory modality. Using the auditory version of the attentional blink paradigm, the present study investigates whether certain auditory objects relating to musical and perceptual expertise could have an impact on the transient costs of selective attention. In this study, expert cellists and novice participants were asked to first identify a target sound, and then to detect instrumental timbres of cello or organ, or human voice as a second target in a rapid auditory stream. The results showed moderate evidence against the attentional blink effect for voices independent of participants’ musical expertise. Experts outperformed novices in their overall accuracy levels of target identification and detection, reflecting a clear benefit of musical expertise. Importantly, the musicianship advantage disappeared when the human voice served as the second target in the stream. The results are discussed in terms of stimulus salience, the advantage of voice processing, as well as perceptual and musical expertise in relation to attention and working memory performance.

Drawing inspiration from the findings in the visual modality, where images of common expertise objects (i.e., human faces) were observed to reduce one’s susceptibility to the visual AB phenomenon (Awh et al., 2004; Landau & Bentin, 2008), and the proposal of unifying coding mechanisms for face and voice processing (Yovel & Belin, 2013), Paper I explored whether musical expertise and common perceptual expertise with human voices and instrumental timbres relating to one’s musical expertise can modulate the temporal costs of selective auditory attention reflected in the auditory AB task.

5. Research summary

The introductory section of Paper I sets the ground for the central themes of this thesis by introducing some of the key findings in the literature that make us suspect that human voices and, by extension, the timbres that share perceptual similarity to voices, can potentially lead to the alleviation or elimination of the AB effect. Accordingly, the following two research questions were posed in Paper I:

- RQ1: Are human voices less susceptible to auditory AB effect than instrumental tones?
- RQ2: If so, does it extend to instruments sharing perceptual similarity to human voice (i.e., cello tones)?

Based on some of the existing evidence for a general musician advantage in the auditory AB task, another aspect that the Paper I set out to explore was the modulations of auditory AB through musical expertise in general, and through the interactions with the objects relating to an expert's domain of expertise. Hence, we also sought to find answers to the following research questions:

- RQ3: Do expert cellists show an overall attenuation of the auditory AB as compared with novices due to their extensive auditory training with tones?
- RQ4: Do expert cellists show less auditory AB to the musical timbres associated with their principal instrument (i.e., cello tones) compared to timbres associated with the instruments that they have not been trained on (i.e., organ tones)?

As a secondary goal, we explored the relationship between the participant-factors of musicality (as measured with Gold-MSI), WM span (measured through LNS), and AB size (measured with maximal AB) of each individual.

5.1.2 Key findings

Results from Paper I showed that human voices as T2s, as compared to the cello and organ tones, were least likely to show an auditory AB effect. To answer RQ1, human voices were indeed less susceptible to the auditory AB effect than the instrumental tones that we have tested here. Importantly, this was the case in both the expert cellist and in the non-musician participant groups. The results also provided a moderate amount of evidence in support of the hypothesis that the conditional accuracy of T2 is not dependent on the lag when T2 is a voice. This result is conclusive (Dienes, 2014) in suggesting that voices as T2s are unlikely to show an auditory AB effect. Combining these three key findings for the human voice we can state that human voices were not only least likely to show an auditory AB effect compared to the instrumental tones, independently from one's musical expertise (i.e., regardless of being an expert cellist or a non-musician), but also unlikely to show an AB effect (based on the Bayesian support for the null), at least within the parameters of this experiment.

For cello tones, we observed only anecdotal evidence favoring the null hypothesis over the model with lag. Thus, the results were inconclusive regarding whether or not cello tones were susceptible to an AB effect. The results showed that participant group explained our data the best when T2 was a cello tone. Under this condition, the expert cellists had higher conditional T2 accuracy than the non-musicians. Taken all together, unlike the voices, there was not enough evidence to suggest that the cello targets are not susceptible to auditory AB effect. In order to claim an extension of lag-independence to the cello condition, we would have expected larger (and conclusive) evidence in support of the lag-independence hypothesis. Thus, more data are needed to make further inferences about the auditory AB effect for cello tones. Nevertheless, given that selectively attending to cello tones did differ with musical expertise, the overall attentional benefits observed for the voice T2s did not seem to extend to the cello T2s.

Regarding the research question concerning the overall attenuation of an auditory AB among expert cellists as compared to the novices, while the frequentists analysis indicated no reliable auditory AB in either group, with the Bayesian analysis we observed moderate evidence in support of no auditory AB effect among the cellists while for the novices the evidence was only anecdotal. This means that one can conclusively state that the auditory AB effect was unlikely to occur among the expert cellists, given our data. To make strong claims regarding the AB effects for the novice group, we would simply need more data. When pitting one group against the other, however, the auditory AB was less likely to occur among the expert cellists than the novices. On average, the maximal AB scores of the expert cellists were also smaller than that of the novices, remarkably, in all but the voice condition, where the maximal AB scores between the two groups no longer differed (moderate evidence for no difference).

The results of the comparison of the likelihood of an AB effect for cello T2s and for organ T2s among the expert cellists showed only anecdotal evidence in support of a difference. The performances of the cellists in general reached ceiling level across all conditions, and these ceiling level performances might have hindered the observation of any meaningful contrasts among the cellists' performance in the auditory AB task. We suggested that making adjustments to the task to make it more sensitive for experts could be useful, should one wish to study auditory AB among experts in the future.

All in all, the stimulus-related findings of this study lend support to the hypothesis that human voice T2s are likely to escape the auditory AB phenomenon, independently from the musical expertise of the participants.

Last but not least, regarding the participant-related factors, overall, participants' musical sophistication scores were negatively correlated with the size of AB when T2 was a cello tone (moderate evidence for H1). Importantly, when T2 was a human voice, musical sophistication was more likely to be not correlated with the size of the AB (strong evidence for H0 over H1). Evidence for the correlations between LNS and size of the AB were inconclusive.

5.1.3 Open questions

Since the evidence for the lag-independence hypothesis was inconclusive for the cello and organ T2s, Paper I could not provide answers regarding the likelihood of observing an auditory AB effect for cello and organ T2s. Thus, more data are needed to make inferences regarding the modulations of the AB effect through these instrumental tones. Despite the conclusive evidence for the modulations of AB through human voice T2s, with the remarkable finding that voices do not seem to be affected by the ‘blink’ phenomenon in either group, an even stronger demonstration of this would be if we could also provide conclusive evidence with regards to this phenomenon in the control conditions. As stated above, the auditory AB effects among the novices (non-musicians) in general were unclear. This puts the auditory AB phenomenon for ‘musical’ tones into question. We followed some of these traces in Paper III. In terms of the participant-related factors, the association between WM and AB size was also unresolved. Since this could be also simply related to recognition in a RSAP task, the relationship between WM and sound recognition was further explored in Paper II.

In Paper I, we discussed, among other factors, long stimulus duration/slow presentation rate as well as task difficulty as alternative explanations of the results, which Paper I could not rule out. In the literature, the effect of stimulus parameters on the auditory AB phenomenon is not thoroughly investigated. The fact that little is known about the stimulus-related effects makes not only designing an auditory AB study, but also the interpretation of the findings (separately from the potential effects of the stimulus factors) much more difficult. Papers II and III aimed to help in resolving this issue.

5.2 Paper II

5.2.1 Abstract

Two experiments were conducted to test the effects of sound duration and timbre on auditory recognition using a rapid serial auditory presentation paradigm. Participants listened to a rapid stream of very brief sounds ranging from 30 to 150 milliseconds and were tested on their ability to distinguish the presence from the absence of a target sound selected from various sound sources placed amongst the distracters. Experiment 1a established that brief exposure to stimuli (60 to 150 milliseconds) does not necessarily equate impaired recognition. In Experiment 1b we found evidence that 30 milliseconds of exposure to the stimuli significantly impairs recognition of single auditory targets, but the recognition for voice and sine tone targets impaired the least, suggesting that the lower limit required for successful recognition could be lower than 30 ms for voice and sine tone targets. Critically, the effect of sound duration on recognition completely disappeared when differences in musical sophistication were controlled for. Our behavioural results extend the studies oriented to understand the processing of brief timbres under temporal constraint by suggesting that the musical sophistication may play a larger role than previously thought. These results can also provide a working hypothesis

for future research, namely, that underlying neural mechanisms for the processing of various sound sources may have different temporal constraints.

Within the general scope of this thesis, in Paper II, the open question from Paper I concerning the influence of the experimental parameters and design choices in the RSAP stream were addressed. Importantly, these influences were tested outside of the AB paradigm in order to isolate them from the factors related to AB. Here we also explored the role of musicality more broadly (i.e., not as a category but as a quantity).

This paper starts with a comprehensive review of the literature on minimum duration thresholds for auditory recognition and on recognizing timbre in brief sound durations. Though I was fascinated and inspired by the observations in the literature demonstrating human auditory recognition capacity, in particular for voices, under remarkably short duration thresholds, I argued that some of the contextual factors in the design of previous investigations (e.g., sound presentation in isolation, carry-over effects) might have played a role in these findings.

Aiming to test the effects of sound duration and sound source on auditory recognition in a RSAP of brief sounds without prior practice/training in the task or the stimuli, we asked two fundamental questions in the two experiments of Paper II:

- RQ1: What is the impact of brief sound durations and timbre on auditory recognition in a rapidly changing auditory environment?
- RQ2: When searching among environment sounds, are certain timbres (e.g., voices) recognized at shorter durations than other sounds (e.g., instrument tones, pure tones, etc.)?

Secondary aim of this paper was to investigate the question of whether differences in musical sophistication (Experiment 1a and 1b) and WM (Experiment 1b) could be associated with the individuals' auditory recognition performance.

5.2.2 Key findings

Regarding RQ1, for the durations between 60 and 150 ms, we found very strong evidence against the model with duration, suggesting that stimulus duration is unlikely to impact recognition sensitivity for a single target presented in RSAP (Experiment 1a). Interestingly, however, in Experiment 1b, when replacing 150 ms sounds with 30 ms in the experimental design, we found extreme evidence favoring the effect of duration over the null model. Duration also explained the recognition sensitivity results best in this experiment compared to other predictors. This finding was due to the decline in recognition in the 30 ms condition, compared to all other durations tested in this experiment. Interestingly, when musical sophistication (measured via Gold-MSI general musical sophistication scores) was added as a covariate, the model with stimulus duration failed to reject the null hypothesis. Together, the results from Paper

5. Research summary

II suggest that stimulus duration only matters for the recognition of extremely brief (i.e., 30 ms) sounds in RSAP, but only when the participants' musical sophistication is not controlled for.

Regarding RQ2, we observed that the recognition of certain timbres was impaired more than others' in the shortest duration that we tested. In particular, in the 30 ms condition, the decline for the recognition of bell and cello tones was sharper than that for voices and sine tones, when presented as a single target among environmental sounds in RSAP.

Concerning the participant-related factors, in Experiment 1a, with traditional statistics, testing the correlations between recognition sensitivity and musical sophistication, we failed to reject the null hypothesis. Further Bayesian analyses showed that the evidence in support of H0 was conclusive (moderate evidence), suggesting that the participants' musical sophistication and recognition sensitivity scores in this study was more likely not to be correlated. In Experiment 1b, again with the traditional approach, we observed a positive correlation between participants' musical sophistication and recognition sensitivity. Further Bayesian analysis indicated a moderate support for this relationship. Musical sophistication scores predicted recognition sensitivity and when musical sophistication was controlled for, the effect of duration no longer reached statistical significance.

Classical correlations with WM scores and recognition sensitivity have failed to reject the null hypothesis. Indeed, Bayesian evidence was conclusive for H0, indicating that digit span and recognition sensitivity scores are more likely not to be related (than to be related).

In addition to the recognition sensitivity index (d-prime), the dependent measures also included the RTs and the subjective certainty ratings to provide additional insight into the recognition processes. The findings obtained from these additional measures are reported and discussed in Paper II. I will not go into the details of those findings here, as the main focus of this study (as illustrated by the RQs) was the results obtained from the recognition sensitivity measure.

5.2.3 Open questions

The findings of Paper II allowed us to narrow down the duration conditions to be included in the auditory AB experiment in Paper III. However, at this point we still did not know whether the auditory AB effect would be influenced by stimuli presented for 30 ms (where single target recognition varied for different timbres) and for 90 ms (where single target recognition was highly similar across different timbres), and whether these influences on the AB effect for various T2s would also differ.

This study has also given rise to several new questions, some of which can be summarized as:

- Is it possible that the enhanced recognition of human voices is due to long-term familiarity and deeper encoding? This question was partly addressed

in Paper III through inclusion of another highly familiar sound to human beings, that is, dog barks.

- What is it that would lead to a better recognition for sine tones than bell tones? Contextual and acoustical differences are discussed among the potential explanations for this result in Paper II. Both these categories have then served as T1s in Paper III so that the T1 task difficulty is more controlled.
- Is it possible that the underlying mechanisms for the processing of different timbres have different temporal constraints? This question can be further studied using neuroimaging techniques.

5.3 Paper III

5.3.1 Abstract

Attentional selection of a second target in a rapid stream of stimuli embedding two targets tends to be briefly impaired when two targets are presented in close temporal proximity. This brief impairment is known as an attentional blink (AB) effect. The present study aims to examine the link between locus coeruleus activity (as measured indirectly through pupil dilations) and the stimulus related factors during an auditory AB task. Previous research has introduced pupillometry as a useful marker of cognitive processing and effort in the visual AB paradigm. Our main objective was to explore mental effort associated with processing and attentional selection of human voices since there are reasons to believe that this class of stimuli escapes AB. Two target sounds (T1 and T2) were embedded in a rapid serial auditory presentation of environmental sounds with a short (Lag 3) or long lag (Lag 9). Participants were to first identify T1 (bell or sine tone) and then to detect T2 (present or absent). Stimuli had durations of either 30 or 90 milliseconds. The T2 varied in category: human voice, cello, or dog sound. Results suggest that the interplay of stimulus factors is critical for both the behavioural target accuracy and the pupil response, and provides support for the hypothesis that human voice is the least likely to show an auditory AB (only in the 90 milliseconds condition). For the other stimuli, accuracy for T2 was significantly worse at Lag 3 than at Lag 9 in the 90 millisecond condition, suggesting the presence of an auditory AB. When AB occurred, we observed larger pupil dilations at Lag 9 compared to Lag 3, reflecting increased attentional processing or effort when targets were temporally distant. Larger pupil dilations for Lag 9 were observed even when our behavioural data did not indicate a reliable auditory AB effect. Taken together, our findings show that the auditory AB effect (or lack thereof) is reflected in pupillary fluctuations, suggesting a potential involvement of the locus coeruleus-noradrenaline system in the AB phenomenon.

Paper III starts with an overview of the most relevant theoretical accounts and investigations, which help linking the various aspects of this study (i.e., pupillometry methodology, AB phenomenon, and LC-NA system), and an introduction

of our starting point for this study. In particular, linking the LC-NA system's activity to the AB phenomenon, we introduce the neurocomputational account of AB (Nieuwenhuis et al., 2005), and the pupillary investigations of AB in the visual modality as well as the evidence on the coupling of pupil response and the LC-NA system (Joshi et al., 2016).

Based on the findings of Paper I, Paper III proposed that human voices may be less susceptible to the auditory AB effect and also followed up on some of the open questions from our previous investigations. Anchored in an auditory AB task combined with a pupillometry methodology, Paper III addressed how the eye's pupil, as an indirect index of the LC-NA system's activity, can inform us about the amount of mental effort associated with the attentional selection of human voices, as compared with other target sounds and under two different stimulus durations.

As mentioned above, the role of stimulus duration was a potential confound on the AB results that Paper I could not rule out. In particular, we discussed in Paper I that the long stimulus duration/slow presentation rate might have led to the observation of no reliable AB in Paper I in general. The experiment reported in Paper III was therefore designed to explore the auditory AB effects in two shorter durations/faster presentation rates. Paper III also contrasted target accuracies and pupillary responses when attending to human voice T2s as well as cello and dog T2s in the auditory AB task. Since the evidence for lag-independence for cello tones was inconclusive in Paper I, it was not possible to make any further claims (other than the group differences) about an auditory AB for T2s that are perceptually similar to human voices. We therefore wanted to further investigate the auditory AB for cello tones. Moreover, as the human voices were contrasted only with the instrumental tones, based on the results of Paper I, it was not possible to talk about whether the voice advantage we presented was unique to human voices, nor was it possible to explore the familiarity aspects discussed in Paper II. Therefore, Paper III introduced a new T2 category (i.e., dog sounds) which is both familiar to human listeners and also a biological sound, as in the case of human voices.

Additionally, we investigated the relationship between musicality, impulsivity, AB size, and mean pupillary responses for each individual.

5.3.2 Key findings

The behavioural results from Paper III partially replicated the main human voice finding from Paper I, showing that the auditory AB effect was the least likely for human voices (as compared with cello and dog sounds) in the 90 ms, but not in the 30 ms condition. The likelihood of human voices (compared to other T2s) to escape the temporal deficits of selective attention reflected in the AB phenomenon seem to be dependent on the stimulus duration/presentation rate (Note that this is not the simple main effect of duration, but the interaction effect). For other T2s, too, under extremely brief durations (i.e., 30 ms) the auditory AB was not observed. We discussed, based on the AB literature, why the 90 ms condition might be the 'sweet-spot' for an AB effect to occur.

Similarly, task-related pupil dilations were also the least likely to be smaller when T2 appeared at Lag 3 than at Lag 9 for human voices in the 90 ms condition (as compared to the other T2s). In addition, the pupil dilations for human voices in the 90 ms condition also had strong support for the null hypothesis (as compared with the alternative hypothesis), suggesting conclusively that the amount of mental effort is unlikely to be smaller at Lag 3 than Lag 9 for human voice T2s in the 90 ms condition.

In Paper III, reliable auditory AB effects were observed for both cello and dog sounds in the 90 ms condition. Critically, unlike Paper I, Paper III conclusively demonstrated evidence for the presence of an auditory AB (in 90 ms) for musical sounds that share acoustic similarity to human voices, that is, the cello (strong evidence) and for biological sounds that share the familiarity aspect, that is, the dog T2s (moderate evidence). In those conditions where the AB effect was present, the pupillary results indicated larger pupil dilations at the long lag, where the T2 deficit is expected to recover. To summarise, paralleling the behavioural results, participants' pupillary dilations (in 90 ms) were larger for T2 presentations at long compared to at short lag following T1, when selectively attending to cello and dog T2s but not to human voices.

With regards to the participant-related factors, classical correlational tests with musicality (all Gold-MSI factors, including musical sophistication) and AB size or mean pupil dilations in each duration condition have failed to reject the null hypothesis. Bayesian analyses revealed moderate evidence in favour of H₀, suggesting that musicality was more likely to be not related (than to be related) to AB size or mean pupillary dilations. Similarly, testing correlations between impulsivity (functional and dysfunctional) and AB size or mean pupil dilations with traditional statistics, we have failed to reject the null hypothesis. Bayesian correlation with functional impulsivity suggested that functional impulsivity was more likely to be not associated with AB size, while the evidence for the correlation with dysfunctional component of impulsivity was only anecdotal. Correlations with pupil responses and impulsivity were also conclusive to suggest that they are more likely to be not related, except for the dysfunctional impulsivity and mean pupillary dilations. This correlation received only anecdotal support for the null hypothesis.

Together, the main findings from Paper III add support to our claim that pupil responses, as an indirect measure of LC activity, can reveal the amount of mental effort in an auditory AB task. These findings also suggest that pupillometry methodology combined with the AB task can have important theoretical implications through offering new ways of studying the auditory AB.

5.3.3 Open questions

While Paper III extended the findings of the experiment in Paper I based on the neurophysiological data that showed the amount of cognitive processing during attentional selection of the stimuli in the auditory AB task, the timing of the auditory AB was not explored within the scope of this study.

5. Research summary

The aspects related to dysfunctional impulsivity, AB size and mental effort in two duration conditions have remained unclear from the present investigation. To make further inferences about the associations between these, we need more data.

Chapter 6

General discussion

“At the end of the day, at the end of the week, at the end of my life, I want to say that I contributed more than I criticized.”

— Brené Brown

This chapter provides a summary and discussion of the findings of the papers as a synthesis, in relation to the issues presented in the two background chapters of this thesis. I will also discuss strengths and limitations of the work presented here, and offer some directions for future research.

6.1 Introduction

Despite more than thirty years of researchers’ efforts to learn about temporal selective attention, relatively few attempts have been made to study this topic in the auditory modality. As we discovered earlier in Chapter 3, some researchers have suggested that temporal limitations as reflected in the AB phenomenon result from a central, amodal limitation (e.g., Arnell & Jolicoeur 1999; Arnell & Larson 2002; Goddard & Slawinski 1999; Shen & Alain 2010), while others have argued for a modality-specific limitation instead (e.g., Duncan et al. 1997; Potter et al. 1998; Soto-Faraco & Spence 2002; Martens et al. 2009). This long-lasting yet unresolved debate, together with surprisingly few investigations of the auditory AB, illustrates that the nature and the scope of the attentional limitations in the auditory modality are still far from being fully understood.

The overarching goal of this thesis was to elucidate whether it is possible to modulate temporal limits of auditory attention and perception through stimulus- and participant-related factors. This would further our understanding of a potential overlap between auditory and visual modalities. Earlier I have described that recent findings from the visual modality reflect a crucial paradigm shift for the theoretical understanding of the AB phenomenon. These findings have stimulated interest in exploring the modulations of the temporal dynamics of human attention through various experimental manipulations and in studying specific groups of people who either show reduced/absent or larger/longer AB. Accumulating evidence suggests that it is possible to overcome the limits of visual attention. Less explored are the modulations of these limits in the auditory modality. Since this is a rather understudied subject, we currently know little about the ways in which these limits may differ (or overlap) with the findings obtained in the visual modality and how to modulate them in the auditory domain.

In Chapter 2, I have reviewed evidence demonstrating that human voices may play a key role in recruiting attentional mechanisms. By contrasting human voices with stimuli that share certain similarities with voices, the investigations in this thesis sought to answer whether human voices can escape the temporal limits of attention and perception. Further manipulations of duration and lag allowed us to understand the temporal aspects of this topic. Based on some of the recent findings suggesting that individuals largely differ in terms of the extent to which they show a visual AB effect (see Chapter 3), next, I asked what are some of the participant factors that can modulate these limits in the auditory modality. The factors of musicality, expertise, WM, and impulsivity were explored. Finally, I asked a methodological question regarding the use of pupillometry as a tool in studying these limits to further understand the differences in terms of mental effort during a behavioural phenomenon.

6.2 Summary and discussion of research contributions

Three main findings arose from the investigations in this thesis:

1. **Sound category can modulate the temporal limits of auditory attention and perception, but its interplay with stimulus duration is critical.**

In this thesis, sound category was shown to modulate the limits of attention and perception. More precisely, we showed that human voices (when presented as T2s) can escape the temporal limits of attention, even when compared to musical timbres that they share acoustic similarity with (i.e., cello), musical timbres that are generally assumed to be familiar but differ acoustically from voices (i.e., organ), and biological sounds that are highly familiar to humans (i.e., dog sounds). Even when non-significant findings were obtained for an auditory AB effect in all stimulus conditions, the evidence was conclusive to suggest that voices were least likely to show an auditory AB effect compared to other sounds (Paper I). The exception to this was when the sounds were presented for 30 ms (Paper III).

In sum, these findings suggest a modulation of attention allocation for human voices in the AB paradigm, at least in some duration conditions. These findings accord with the idea that certain stimuli can be less susceptible to AB because of their higher priority or salience (Shapiro et al., 1997a). This could be further explained with the evolutionary and/or the neuroscientific perspectives on the significance of voices that are covered extensively in Chapter 2. Nevertheless, I would be cautious to interpret these results to imply that voices will always escape auditory AB effects. These data, together with the literature I reviewed in Chapter 3 on modulations of AB through stimulus factors, mainly suggests that temporal limits of attention can be modulated in various ways, but the context (i.e., other experimental parameters) matters. For example, the findings

obtained with face stimuli pointed to the conclusion that faces *can, but do not always*, escape the AB. Indeed, our findings in Paper III, too, indicated that when each sound was presented for 30 ms, the voice advantage on overcoming the temporal limits of attention was no longer observed.

An alternative interpretation of these results could be that low-level differences are the underlying cause of how voices may be able to overcome these restrictions. This will be discussed further in Section 6.3. Although it is not possible to fully eliminate this explanation without conducting new experiments designed directly to test this question, it is unlikely that these results are solely driven by low-level factors. There are several grounds for this argument. One reason for this is the fact that both the distracters and the other T2s were perceptually rich. Second, we did not obtain conclusive evidence in support of the hypothesis that the voice advantage observed in the AB task would extend to sounds that shared acoustic similarity with voices. However, in Paper II, where the aim was to investigate single target processing separately from AB, we did observe that the recognition of cello sounds suffered more from being shortened to 30 ms than that of voices. This, when considered together with the findings from Paper III, is surprising, since the 30 ms condition was the only condition where the voice advantage hypothesis in the AB task was not supported.

The findings on the modulations through stimulus duration manipulations were more mixed. Although the main effect of duration was not always supported, duration generally seemed to have an interaction with other experimental manipulations. For example, a significant auditory AB effect was observed only for cello and dog T2s in the 90 ms duration (Paper III), while we failed to reject the null hypothesis for an AB effect in the 150 ms duration (Paper I). Single target recognition, however, was only impaired in the 30 ms duration (Paper II). Together, these results can be interpreted to suggest that the modulation of AB and single target processing within RSAP is possible through an interplay between stimulus type and stimulus duration. However, why temporal attentional limitations should be more restricted in the 90 ms while single target processing seems more restricted in the 30 ms remains to be resolved. Potential explanations for these are discussed in the individual papers in detail.

2. Modulation of the temporal limits of auditory attention and perception through participant-related factors needs further investigation.

The present investigations showed mixed and/or inconclusive evidence regarding the associations of musicality/expertise, WM, and impulsivity with task performance. Overall, these results are in line with the findings in the literature in their inconclusiveness. This further suggests that this is a rather complex issue that seems to be dependent on the specifics of the experimental design and/or how these concepts were measured.

In the investigations in this thesis, first, contradicting findings emerged regarding the associations between musical sophistication and target recognition within a rapid stream of brief sounds in the two investigations in Paper II. More precisely, while Experiment 1a suggested that these were likely not correlated, Experiment 1b indicated an opposite pattern. It is possible that the contrast between these results was, at least partly, driven by the performances in the 30 ms duration condition. This is because in Experiment 1b, controlling for the differences in participants' musical sophistication influenced our findings regarding the effect of duration. Prior to this control, the effect of duration was shown to be due to the decline in recognition in the 30 ms condition. This suggests that the individual differences in musicality seem to be particularly relevant for processing sounds at extremely brief duration thresholds. The findings of Papers I and III were also mixed. In Paper III, musical sophistication was found more likely to be unrelated to AB size (than to be related). In Paper I, we observed that the support for/against these correlations differed under different T2 conditions and between expert cellists and non-musicians. Together, these findings reveal a complex pattern for the potential modulations through musical sophistication and/or expertise which seem to be influenced in some ways by the stimulus-related factors or differences in task demands.

Regarding WM findings, in Paper I, the evidence was not conclusive to make inferences regarding the link between participants' WM span and their AB size. In Paper II, the evidence was conclusive to suggest that WM span was not related to single target recognition in a RSAP task. This difference could be due to many reasons. Some of these include differences between the tasks, participants, or measurements. Inconclusive evidence in Paper I, together with the mixed findings in the literature regarding the associations between WM and AB size (as pointed out in Subsection 3.4.3), reveal that further research is needed to understand whether WM can be related to differences in the temporal limits of attention.

Finally, according to our findings in Paper III, the functional component of impulsivity did not seem to be associated with AB size, but there was not enough evidence to draw any conclusions about the associations with the dysfunctional component of impulsivity. These findings seem partly in contrast with the findings from an earlier study, in which it was reported that non-blinkers were more functionally impulsive than blinkers (Troche & Rammsayer, 2013). From this finding, one would have expected to observe a negative correlation between functional impulsivity and AB size. The differences between these studies (e.g., modality difference, the comparison of non-blinkers and blinkers versus the individual differences approach) could potentially be responsible for the differences in the functional impulsivity findings. However, the findings of these two studies are not sufficient to draw firm conclusions. Moreover, since the evidence we obtained was not conclusive, any inferences on our dysfunctional impulsivity results would only be speculative. However, it is safe to state that these findings

indicate that not only the functional but also dysfunctional component deserves further exploration. All in all, these results highlight that future AB studies should focus on exploring impulsivity not as a single construct, but together with its components.

3. Pupillometry is a useful method to reveal the mental effort associated with the allocation of temporal selective attention to brief sounds.

Since AB, in itself, is a behavioural effect, accuracy results alone were not enough to inform us on how intensely people were attending to report the targets. Pupil data provided additional insights into what was happening ‘behind the scenes’ within each experimental condition. To make my point even clearer, I would like to borrow an analogy of a swimming duck from Stanford University ¹. By only observing a duck floating peacefully on the surface (behaviour), we could easily miss the effort (here in the physical sense) that goes into paddling continuously underneath the water. Through pupillometry, we were able to evaluate and compare the amount of mental effort (or attentional resources) required for detecting (and failing to detect) brief auditory targets during an auditory AB task. Since the AB constitutes a failure in attentional allocation, there should be less pupillary dilation when participants show an AB compared with when they detect T2.

More specifically, in Paper III, we showed that target detection is related to increased pupil dilations, reflecting larger mental effort, similar to the previous reports in the vision literature (Privitera et al., 2010). We also observed that the human voice T2s (in the 90 ms condition) were the least likely to have this increase in pupil dilations. This finding suggests that selectively attending to human voices may require less mental effort than other auditory stimuli. Finally, when T2 was a dog sound in the 30 ms duration condition, behaviourally, the participants were least likely show an AB effect. However, the intensity of the underlying cognitive processes, as reflected in their pupil dilations, was telling another story (similar to the duck analogy). Here we observed that for the dog T2s, there was a significant increase in the pupil response at the later lag where temporal deficits associated with the AB effect are supposed to recover. We argued that this result can be interpreted in a couple of ways: 1) It could be that detecting dog T2s (or the expectation of them) is in general more mentally taxing than the other T2s. 2) It could also be that detecting briefly presented dog sounds causes more uncertainty or surprise than other target types (possibly due to their acoustic qualities). This is because uncertainty and surprise have also been shown to dilate the pupils (e.g., Nassar et al. 2012; Preuschoff et al. 2011).

¹This analogy was originally used to refer to students’ struggles in what is known as the Stanford Duck Syndrome. Sun, T. (2018). Duck syndrome and a culture of misery. *The Stanford Daily*. January 31, 2018. <https://stanforddaily.com/2018/01/31/duck-syndrome-and-a-culture-of-misery/> (Last visited May 9, 2022).

Together, these results show that pupillometry is a useful tool to complement findings from a fast-occurring behavioural phenomenon such as the AB, not only in the visual (Wierda et al., 2012), but also in the auditory modality. Although our findings do not directly speak to the underlying causal relations between LC activity and temporal selective attention, this method offered an indirect index to the LC-NA system's activity. According to the present findings (as the observed modulations suggests that AB is not a hard-wired limitation), if LC plays a role in the AB phenomenon as claimed in the neurocomputational model of AB (Nieuwenhuis et al., 2005), then either the magnitude of the NA release (De Martino et al., 2008), or the potentiating requirement (Nieuwenhuis et al., 2005), or the timing of the LC refractory period should be different for high priority stimuli such as human voices compared with other stimuli.

To conclude, the findings presented above are consistent with Shen & Alain's (2012) conclusion that the temporal limits of human cognition do not seem to be rigid, or at least, some degree of flexibility seems to exist to overcome these restrictions, also in the auditory modality. This is also in line with recent findings in the visual modality on the modulations of attentional restrictions through manipulations of stimuli and tasks, and differences due to various participant factors, as mentioned in Section 3.4. These findings critically question what clearly has a grounding in the earlier theories of the AB phenomenon. One useful approach to the problem of bringing together the past and present of AB studies could be to make a distinction between 'strong' and 'weak' constraints.

6.3 Limitations and Strengths

6.3.1 Limitations

A limitation of the investigations presented in this thesis concerns the time-course of attention. It is plausible that a deficit in attending to targets may have exhibited itself at an earlier or later lag, beyond the lags included in our experimental designs. Since this thesis aimed to explore several aspects involved in modulating temporal attention and perception, each study design was rather complex. The flip side of this was that the levels within each experimental condition had to be restricted. In simpler designs, it could have been possible to include a full range of lags (e.g., from Lag 1 through Lag 9). As mentioned elsewhere (see discussion in Paper III), this compromise limited the opportunity to explore the full time-course of temporal selective attention.

Next, the AB task in the present investigations included a task-switch element (i.e., a switch in perceptual set between T1 and T2 tasks). According to Potter and colleagues (1998), the presence of an AB effect in the auditory modality reflects task-switch costs. Based on this argument, one could argue that task-switch explains the AB findings in the investigations in this thesis. If I were to conduct these studies again, I would have controlled for this. Nevertheless, given that the task-switch element remained the same, but the auditory AB effects (or

lack thereof) varied with our experimental manipulations, it seems unlikely that task-switch can fully explain the present findings.

Pertaining to the aspects of sound stimuli, first, a better control of the low-level acoustic aspects of the stimuli would have been ideal. In particular, subjective loudness of stimuli may have differed due to the use of peak normalization method. However, in this instance, using an RMS method would have come at a high cost. Earlier it was shown by Bigand et al. (2011) that the voice advantage only appeared in the RMS normalized set. If I had used this method, a potential voice advantage in overcoming the temporal limits of auditory attention and perception could have been explained by the RMS normalization. Second, as a natural consequence of studying brief sounds, the ecological validity of the sound stimuli used in these investigations was low. In addition, it has been argued that sung vowels and isolated tones are weakly representative of spoken voices and music, respectively (Bigand et al., 2011). On this basis, generalizing the findings by Agus et al. (2012) to everyday sounds was suggested to be difficult. This criticism applies also to the papers in this thesis. Using a wider set of stimuli would be beneficial to increase the representativeness of the sound categories.

Another issue concerns the explorations of the modulations of AB and target processing in RSAP through a selection of factors. It is, of course, possible that the AB magnitude is related to other factors that were not tested (or controlled for) in this dissertation. For example, regarding participant-factors, dispositional affect and personality traits (including but not limited to impulsivity) have been shown to predict visual AB magnitude, as mentioned earlier in Chapter 3. It is plausible that these (or other) factors may be related to the present findings in the auditory modality. Likewise, it is also plausible that factors that we did explore may be related to other factors. An example of this could be that the findings related to musicality might be related to (or derive from) the listeners' motivation or arousal levels during an auditory task. A related issue is the role of measurement choices. For instance, as previously indicated, findings in the literature concerning the association between WM and AB are mixed. These, as well as the WM findings in the present investigations, could be due to the differences between the measurement. Future investigation would benefit from using multiple measures of the same construct.

6.3.2 Strengths

By studying the uncharted territories of a phenomenon that made a huge impact in the field of cognitive psychology, this thesis has addressed several gaps in the literature. These included primarily the understudied topic of temporal limits of attention in the auditory modality, but also the study of voice cognition and musical timbre in a novel context.

Through its background chapters, combining knowledge from various scientific disciplines, approaches, and sensory modalities, this thesis has helped to move the field towards a more unified understanding of human temporal selective attention and perception mechanisms beyond traditional boundaries.

The framework and methodology used in the individual papers have set novel groundwork for future studies aiming to elucidate how we perceive and distribute our attention to brief sounds in time. This can be beneficial for studies of auditory cognition and music perception. As Jensenius (2007) suggests, music perception should be studied together with the limits and capacities of the human auditory system, as the evolutionary process that has developed this system forms the foundation for perception, performance and composition of music.

The methodological strengths of this thesis include the novel use of the RSAP task to investigate AB with complex sound stimuli, the novel application of pupillometry method to investigate AB in the auditory modality, and design choices that prioritized compatibility between experiments. Furthermore, through pairing NHST with the Bayesian approach to statistics, it has provided a far more thorough picture of the findings and pushed the field forward towards higher standards of statistical reporting.

Importantly, this thesis, through offering new insights that challenge the inevitability of the AB in the auditory modality, has contributed to the recent shift in the empirical and theoretical direction of temporal selective attention studies. Furthermore, the topic and the insights of this thesis are connected to many interesting and significant topics. Next, I will offer some of these as directions for future work.

6.4 Future directions

Many aspects of the potential modulations of temporal auditory attention and perception are yet to be explored. Future explorations could include manipulations of various other qualities of sound signals, complementing the stimulus-related factors that I have explored in this thesis. Similarly, future explorations can benefit from identifying other potential participant-factors that may modulate these mechanisms. This could help to elucidate the individual differences in task performance observed in recent AB studies. An important challenge for this line of work is exploring how these factors may interact with other factors (e.g., intelligence), which may not always be possible to control for. Due to the correlational nature of such investigations, one should be careful with the conclusions drawn.

Several exciting directions for future research exist also within the topic of human voice cognition. Future studies can explore whether there could be a human voice advantage in other cognitive phenomena as well (e.g., psychological refractory period, repetition deafness, change deafness etc.). Moreover, direct comparative studies of the temporal aspects of face and voice processing have also a great potential to further our current understanding of voice and face cognition and how these processes are integrated. Future studies can benefit greatly from a multi-modal approach to study the speed and accuracy of the integration of these processes both in healthy and in clinical populations.

Clinical extensions of the topic explored in this thesis are, in general, plentiful.

It could be particularly valuable to study the temporal dynamics of voice cognition and the human voice advantage in autism spectrum disorder, schizophrenia, and dementia. In future work, I hope to address such clinical applications.

Avenues of research that address the time-course of vocal emotion processing are also likely to be fruitful. Also of interest is studying brief vocal emotion processing in relation to gender-emotion stereotypes.

Many of these future directions have both theoretical and societal significance. Further understanding of the extensions and applications of the topic explored in this thesis would therefore be a worthy goal for the future.

My hope is that the contributions made in this thesis have informed us on some of the missing aspects of the literature within AB studies and voice cognition, and will inspire future research. I look forward to the next decades of research on temporal auditory attention and voice processing.

Bibliography

- Agres, K. R. & Krumhansl, C. L. (2008). Musical change deafness: The inability to detect change in a non-speech auditory domain. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Volume 30.
- Agus, T. R., Paquette, S., Suied, C., Pressnitzer, D. & Belin, P. (2017). Voice selectivity in the temporal voice area despite matched low-level acoustic cues. *Scientific Reports*, 7(1), 1–7.
- Agus, T. R., Suied, C., Thorpe, S. J. & Pressnitzer, D. (2012). Fast recognition of musical sounds based on timbre. *The Journal of the Acoustical Society of America*, 131(5), 4124–4133.
- Alnæs, D., Sneve, M. H., Espeseth, T., Endestad, T., van de Pavert, S. H. P. & Laeng, B. (2014). Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. *Journal of Vision*, 14(4), 1–20.
- Altman, D. G. & Bland, J. M. (1995). Statistics notes: Absence of evidence is not evidence of absence. *BMJ*, 311(7003), 485.
- Andics, A. & Faragó, T. (2018). Voice perception across species. In *The Oxford Handbook of Voice Perception* (pp. 363–392). Oxford University Press.
- Arnell, K. M. & Jenkins, R. (2004). Revisiting within-modality and cross-modality attentional blinks: Effects of target—distractor similarity. *Perception & Psychophysics*, 66(7), 1147–1161.
- Arnell, K. M. & Jolicoeur, P. (1999). The attentional blink across stimulus modalities: Evidence for central processing limitations. *Journal of Experimental Psychology: Human Perception and Performance*, 25(3), 630.
- Arnell, K. M. & Larson, J. M. (2002). Cross-modality attentional blinks without preparatory task-set switching. *Psychonomic bulletin & Review*, 9(3), 497–506.
- Arnell, K. M., Stokes, K. A., MacLean, M. H. & Gicante, C. (2008). Executive control processes of working memory predict attentional blink magnitude over and above storage capacity. *Psychological Research*, 74(1), 1–11.
- Arnell, K. M. & Stubitz, S. M. (2010). Attentional blink magnitude is predicted by the ability to keep irrelevant material out of working memory. *Psychological Research*, 74(5), 457–467.
- Aronson, A. E. (1985). *Clinical voice disorders: An interdisciplinary approach* (3 Ed.). New York: Thieme Medical Publishers Inc.

Bibliography

- Aston-Jones, G. & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*(1), 403–450.
- Awh, E., Serences, J., Laurey, P., Dhaliwal, H., van der Jagt, T. & Dassonville, P. (2004). Evidence against a central bottleneck during the attentional blink: Multiple channels for configural and featural processing. *Cognitive Psychology*, *48*(1), 95–126.
- Baddeley, A. D. & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation*, Volume 8 (pp. 47–89). Elsevier.
- Baker, C. A., Clemens, J. & Murthy, M. (2019). Acoustic pattern recognition and courtship songs: Insights from insects. *Annual Review of Neuroscience*, *42*, 129–147.
- 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.
- Bayarri, M. J. & Berger, J. O. (2004). The interplay of Bayesian and frequentist analysis. *Statistical Science*, *19*(1), 58–80.
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276.
- Belin, P. (2006). Voice processing in human and non-human primates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1476), 2091–2107.
- Belin, P. (2019). The ‘vocal brain’: Core and extended cerebral networks for voice processing. In *The Oxford Handbook of Voice Perception*. Oxford University Press.
- Belin, P., Bestelmeyer, P. E., Latinus, M. & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology*, *102*(4), 711–725.
- Belin, P., Fecteau, S. & Bedard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences*, *8*(3), 129–135.
- Belin, P., Zatorre, R. J. & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, *13*(1), 17–26.
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P. & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309–312.
- 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*, 2023.

- Bigand, E., Delbe, C., Gérard, Y. & Tillmann, B. (2011). Categorization of extremely brief auditory stimuli: domain-specific or domain-general processes? *PLoS One*, *6*(10), e27024.
- Blacker, K. J. & Curby, K. M. (2016). The attentional blink is attenuated for objects of expertise. *Visual Cognition*, *24*(2), 132–140.
- Bodin, C. & Belin, P. (2020). Exploring the cerebral substrate of voice perception in primate brains. *Philosophical Transactions of the Royal Society B- Biological Sciences*, *375*(1789), 20180386.
- Brandler, S. & Rammsayer, T. H. (2003). Differences in mental abilities between musicians and non-musicians. *Psychology of Music*, *31*(2), 123–138.
- Braun, J. (1998). Vision and attention: The role of training. *Nature*, *393*(6684), 424–425.
- Broadbent, D. E. & Broadbent, M. H. (1987). From detection to identification: Response to multiple targets in rapid serial visual presentation. *Perception & Psychophysics*, *42*(2), 105–113.
- Bruce, V. & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*(3), 305–327.
- Bugos, J. & Mostafa, W. (2011). Musical training enhances information processing speed. *Bulletin of the Council for Research in Music Education*, (187), 7–18.
- Burton, A. M. & Bruce, V. (1992). I recognize your face but i can't remember your name: A simple explanation? *British Journal of Psychology*, *83*(1), 45–60.
- Buss, D. M. (2012). *Evolutionary psychology: The new science of the mind* (4 Ed.). Pearson.
- Carey, D., Rosen, S., Krishnan, S., Pearce, M. T., Shepherd, A., Aydelott, J. & Dick, F. (2015). Generality and specificity in the effects of musical expertise on perception and cognition. *Cognition*, *137*, 81–105.
- Charest, I., Pernet, C. R., Rousselet, G. A., Quiñones, I., Latinus, M., Fillion-Bilodeau, S., Chartrand, J.-P. & Belin, P. (2009). Electrophysiological evidence for an early processing of human voices. *BMC Neuroscience*, *10*(1), 1–11.
- Chartrand, J.-P. & Belin, P. (2006). Superior voice timbre processing in musicians. *Neuroscience Letters*, *405*(3), 164–167.
- Chartrand, J.-P., Peretz, I. & Belin, P. (2008). Auditory recognition expertise and domain specificity. *Brain Research*, *1220*, 191–198.
- Cherry, C. (1953). Some experiments on the recognition of speech with one. *The Journal of the Acoustical Society of America*, *25*(5), 975–979.

Bibliography

- Choi, H., Chang, L.-H., Shibata, K., Sasaki, Y. & Watanabe, T. (2012). Resetting capacity limitations revealed by long-lasting elimination of attentional blink through training. *Proceedings of the National Academy of Sciences*, *109*(30), 12242–12247.
- Chun, M. M. (1997). Types and tokens in visual processing: A double dissociation between the attentional blink and repetition blindness. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(3), 738–755.
- Chun, M. M. & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 109–127.
- Chun, M. M. & Potter, M. C. (2001). The attentional blink and task switching within and across modalities. In *The Limits of Attention: Temporal Constraints in Human Information Processing* (pp. 20–35). Oxford University Press.
- Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, *1*(3), 98–101.
- Cohen, J. (1994). The earth is round ($p < .05$). *The American Psychologist*, *49*(12), 997–1003.
- Cohen, R. A. (2014). *The Neuropsychology of Attention*. Springer.
- Collins, S. (2004). Vocal fighting and flirting: The functions of birdsong. *Nature's music: The science of birdsong* (pp. 39–79).
- Colzato, L. S., Spapé, M. M., Pannebakker, M. M. & Hommel, B. (2007). Working memory and the attentional blink: Blink size is predicted by individual differences in operation span. *Psychonomic Bulletin & Review*, *14*(6), 1051–1057.
- Cross, I. (2008). Musicality and the human capacity for culture. *Musicae Scientiae*, *12*(1_suppl), 147–167.
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29.
- Darwin, C. (1859). *On the origin of species*. London: Murray.
- Dayan, P. & Yu, A. J. (2005). Norepinephrine and neural interrupts. *Advances in Neural Information Processing Systems*, *18*.
- De Martino, B., Strange, B. A. & Dolan, R. J. (2008). Noradrenergic neuromodulation of human attention for emotional and neutral stimuli. *Psychopharmacology*, *197*(1), 127–136.
- DeCasper, A. J. & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*(4448), 1174–1176.

- Denis, D. J. (2003). Alternatives to null hypothesis significance testing. *Theory & Science*, 4(1), 21.
- Dickman, S. J. (1990). Functional and dysfunctional impulsivity: personality and cognitive correlates. *Journal of Personality and Social Psychology*, 58(1), 95.
- Dickman, S. J. (2000). Impulsivity, arousal and attention. *Personality and Individual Differences*, 28(3), 563–581.
- Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. New York: Palgrave Macmillan.
- Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, 6(3), 274–290.
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, 5, 781.
- Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, 72, 78–89.
- Dolar, M. (2006). *A voice and nothing more*. Cambridge, MA: MIT press.
- Dunbar, R. I. (2017). Group size, vocal grooming and the origins of language. *Psychonomic Bulletin & Review*, 24(1), 209–212.
- Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, 87(3), 272–300.
- Duncan, J., Martens, S. & Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature*, 387(6635), 808–810.
- Dux, P. E. & Harris, I. M. (2007). On the failure of distractor inhibition in the attentional blink. *Psychonomic Bulletin & Review*, 14(4), 723–728.
- Dux, P. E. & Marois, R. (2008). Distractor inhibition predicts individual differences in the attentional blink. *PLoS One*, 3(10), e3330.
- Dux, P. E. & Marois, R. (2009). The attentional blink: A review of data and theory. *Attention, Perception, & Psychophysics*, 71(8), 1683–1700.
- Edwards, W., Lindman, H. & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, 70(3), 193.
- Endestad, T., Godøy, R. I., Sneve, M. H., Hagen, T., Bochynska, A. & Laeng, B. (2020). Mental effort when playing, listening, and imagining music in one pianist's eyes and brain. *Frontiers in Human Neuroscience*, 14, 576888.
- Etz, A. & Vandekerckhove, J. (2018). Introduction to Bayesian inference for psychology. *Psychonomic Bulletin & Review*, 25(1), 5–34.

Bibliography

- Faul, F., Erdfelder, E., Lang, A. G. & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.
- Fecteau, S., Armony, J. L., Joanette, Y. & Belin, P. (2004). Is voice processing species-specific in human auditory cortex? an fmri study. *Neuroimage*, *23*(3), 840–848.
- Feinstein, J. S., Stein, M. B., Castillo, G. N. & Paulus, M. P. (2004). From sensory processes to conscious perception. *Consciousness and Cognition*, *13*(2), 323–335.
- Fifer, W. P. & Moon, C. M. (1994). The role of mother's voice in the organization of brain function in the newborn. *Acta Paediatrica*, *83*, 86–93.
- Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America*, *102*(2), 1213–1222.
- Fitch, W. T. & Fritz, J. B. (2006). Rhesus macaques spontaneously perceive formants in conspecific vocalizations. *The Journal of the Acoustical Society of America*, *120*(4), 2132–2141.
- Fujioka, T., Ross, B., Kakigi, R., Pantev, C. & Trainor, L. J. (2006). One year of musical training affects development of auditory cortical-evoked fields in young children. *Brain*, *129*(10), 2593–2608.
- George, E. M. & Coch, D. (2011). Music training and working memory: An ERP study. *Neuropsychologia*, *49*(5), 1083–1094.
- Ghazanfar, A. A. & Rendall, D. (2008). Evolution of human vocal production. *Current Biology*, *18*(11), R457–R460.
- Giesbrecht, B., Sy, J. L. & Lewis, M. K. (2009). Personal names do not always survive the attentional blink: Behavioral evidence for a flexible locus of selection. *Vision Research*, *49*(10), 1378–1388.
- Gigerenzer, G. (1993). The Superego, the Ego, and the Id in statistical reasoning. *A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues* (pp. 311–339).
- Goddard, K. & Slawinski, E. (1999). Modality specific attentional mechanisms govern the attentional blink. *Canadian Acoustics*, *27*, 98–99.
- Goddard, K. M., Slawinski, E. B. & Isaak, M. I. (1998). The auditory attentional blink in a congenitally blind population. *The Journal of the Acoustical Society of America*, *103*(5), 3018–3018.
- Green, C. S. & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*(6939), 534–537.

- Green, D. M. & Swets, J. A. (1966). *Signal detection theory and psychophysics*, Volume 1. New York: Wiley.
- Gunji, A., Koyama, S., Ishii, R., Levy, D., Okamoto, H., Kakigi, R. & Pantev, C. (2003). Magnetoencephalographic study of the cortical activity elicited by human voice. *Neuroscience Letters*, *348*(1), 13–16.
- Hamilton, W. D. (1964). The genetical evolution of social behaviour ii. *Journal of Theoretical Biology*, *7*(1), 17–52.
- Hanley, J. R. & Cohen, G. (2008). Memory for people: Faces, names, and voices. In *Memory in the real world* (3 Ed.) (pp. 107–136). Psychology Press.
- Harlow, L. L., Mulaik, S. A. & Steiger, J. H. (2013). *What if there were no significance tests?* New York: Psychology Press.
- de Heide, R. & Grünwald, P. D. (2020). Why optional stopping can be a problem for Bayesians. *Psychonomic Bulletin & Review*, *28*(3), 795–812.
- Hess, E. H. & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science*, *132*(3423), 349–350.
- Hess, E. H. & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, *143*(3611), 1190–1192.
- Hill, B., Elliott, E. M., Shelton, J. T., Pella, R. D., O’Jile, J. R. & Gouvier, W. D. (2010). Can we improve the clinical assessment of working memory? an evaluation of the Wechsler Adult Intelligence Scale–Third Edition using a working memory criterion construct. *Journal of Clinical and Experimental Neuropsychology*, *32*(3), 315–323.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H. & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.
- Hommel, B. (2020). Dual-task performance: Theoretical analysis and an event-coding account. *Journal of Cognition*, *3*(1), 29.
- Honing, H. (2018). On the biological basis of musicality. *Annals of the New York Academy of Sciences*, *1423*(1), 51–56.
- Isnard, V., Chastres, V., Viaud-Delmon, I. & Suied, C. (2019). The time course of auditory recognition measured with rapid sequences of short natural sounds. *Scientific Reports*, *9*(1), 8005.
- Jackson, M. C. & Raymond, J. E. (2006). The role of attention and familiarity in face identification. *Perception & Psychophysics*, *68*(4), 543–557.
- JASPTeam (2019). Jasp. *Computer software*.

Bibliography

- Jeffreys, H. (1961). *The theory of probability* (3 Ed.). Oxford: Oxford University Press.
- Jensenius, A. R. (2007). *Action-sound: Developing methods and tools to study music-related body movement*. PhD thesis, University of Oslo.
- Jolicoeur, P. (1998). Modulation of the attentional blink by on-line response selection: Evidence from speeded and unspeeded task1 decisions. *Memory & Cognition*, *26*(5), 1014–1032.
- de Jong, P. J., Koster, E. H., van Wees, R. & Martens, S. (2009). Emotional facial expressions and the attentional blink: Attenuated blink for angry and happy faces irrespective of social anxiety. *Cognition and Emotion*, *23*(8), 1640–1652.
- Joshi, S., Li, Y., Kalwani, R. M. & Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, *89*(1), 221–234.
- Kahneman, D. (1973). *Attention and effort*, Volume 1063. Englewood Cliffs, NJ: Prentice Hall.
- Kahneman, D. & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(3756), 1583–1585.
- Kahneman, D. & Beatty, J. (1967). Pupillary responses in a pitch-discrimination task. *Perception & Psychophysics*, *2*(3), 101–105.
- Kanwisher, N. G. (1987). Repetition blindness: Type recognition without token individuation. *Cognition*, *27*(2), 117–143.
- Kisilevsky, B. S., Hains, S. M., Lee, K., Xie, X., Huang, H., Ye, H. H., Zhang, K. & Wang, Z. (2003). Effects of experience on fetal voice recognition. *Psychological Science*, *14*(3), 220–224.
- Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, D.C: American Psychological Association.
- Knudsen, D. P. & Gentner, T. Q. (2010). Mechanisms of song perception in oscine birds. *Brain and Language*, *115*(1), 59–68.
- Koreimann, S., Gula, B. & Vitouch, O. (2014). Inattentional deafness in music. *Psychological Research*, *78*(3), 304–312.
- Koreimann, S., Strauß, S. & Vitouch, O. (2009). Inattentional deafness under dynamic musical conditions. In *Proceedings of the 7th Triennial Conference of the European Society for the Cognitive Sciences of Music* (pp. 246–249). Jyväskylä, Finland: Dept. of Music, University of Jyväskylä.

- Kreiman, J. & Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. Hoboken: John Wiley & Sons.
- Kriegstein, K. V. & Giraud, A. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *NeuroImage*, *22*(2), 948–955.
- Krueger, J. (2001). Null hypothesis significance testing: On the survival of a flawed method. *The American Psychologist*, *56*(1), 16–26.
- Laeng, B. & Alnaes, D. (2019). Pupillometry. In *Eye Movement Research* (pp. 449–502). Springer International Publishing.
- Laeng, B., Ørbo, M., Holmlund, T. & Miozzo, M. (2011). Pupillary Stroop effects. *Cognitive Processing*, *12*(1), 13–21.
- Laeng, B., Sirois, S. & Gredebäck, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science*, *7*(1), 18–27.
- Landau, A. N. & Bentin, S. (2008). Attentional and perceptual factors affecting the attentional blink for faces and objects. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(4), 818–830.
- Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences*, *9*(2), 75–82.
- Lavie, N. & Tsai, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics*, *56*(2), 183–197.
- Lee, M. D. & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.
- Lee, Y., Lu, M. & Ko, H. (2007). Effects of skill training on working memory capacity. *Learning and Instruction*, *17*(3), 336–344.
- van Leeuwen, S., Müller, N. G. & Melloni, L. (2009). Age effects on attentional blink performance in meditation. *Consciousness and Cognition*, *18*(3), 593–599.
- Levy, D. A., Granot, R. & Bentin, S. (2001). Processing specificity for human voice stimuli: Electrophysiological evidence. *Neuroreport*, *12*(12), 2653–2657.
- Levy, D. A., Granot, R. & Bentin, S. (2003). Neural sensitivity to human voices: ERP evidence of task and attentional influences. *Psychophysiology*, *40*(2), 291–305.
- Li, C. R., Chen, S., Lin, W. & Yang, Y. (2005). Attentional blink in adolescents with varying levels of impulsivity. *Journal of Psychiatric Research*, *39*(2), 197–205.

Bibliography

- Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*, Volume 8. Oxford University Press.
- Macdonald, J. S. & Lavie, N. (2011). Visual perceptual load induces inattentional deafness. *Attention, Perception, & Psychophysics*, *73*(6), 1780–1789.
- Mack, A. & Rock, I. (1998). *Inattention blindness: Perception without attention*. Cambridge, MA: MIT Press.
- MacLean, M. H. & Arnell, K. M. (2010). Personality predicts temporal attention costs in the attentional blink paradigm. *Psychonomic Bulletin & Review*, *17*(4), 556–562.
- MacLean, M. H. & Arnell, K. M. (2012). A conceptual and methodological framework for measuring and modulating the attentional blink. *Attention, Perception, & Psychophysics*, *74*(6), 1080–1097.
- MacLean, M. H., Arnell, K. M. & Busseri, M. A. (2010). Dispositional affect predicts temporal attention costs in the attentional blink paradigm. *Cognition and Emotion*, *24*(8), 1431–1438.
- Macmillian, N. A. & Creelman, C. D. (2005). *Detection Theory: A User's Guide* (2 Ed.). New York: Lawrence Erlbaum Associates.
- Maratos, F. A. (2011). Temporal processing of emotional stimuli: The capture and release of attention by angry faces. *Emotion*, *11*(5), 1242.
- Martens, S. & Johnson, A. (2008). Working memory capacity, intelligence, and the magnitude of the attentional blink revisited. *Experimental Brain Research*, *192*(1), 43–52.
- Martens, S., Johnson, A., Bolle, M. & Borst, J. (2009). A quick visual mind can be a slow auditory mind: Individual differences in attentional selection across modalities. *Experimental Psychology*, *56*(1), 33–40.
- Martens, S., Munneke, J., Smid, H. & Johnson, A. (2006). Quick minds don't blink: Electrophysiological correlates of individual differences in attentional selection. *Journal of Cognitive Neuroscience*, *18*(9), 1423–1438.
- Martens, S. & Valchev, N. (2009). Individual differences in the attentional blink: The important role of irrelevant information. *Experimental Psychology*, *56*(1), 18–26.
- Martens, S., Wierda, S. M., Dun, M., de Vries, M. & Smid, H. G. (2015). Musical minds: attentional blink reveals modality-specific restrictions. *PloS One*, *10*(2), e0118294.
- Martens, S. & Wyble, B. (2010). The attentional blink: Past, present, and future of a blind spot in perceptual awareness. *Neuroscience & Biobehavioral Reviews*, *34*(6), 947–957.

- Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition*, 1(1).
- McHugo, M., Olatunji, B. O. & Zald, D. H. (2013). The emotional attentional blink: What we know so far. *Frontiers in Human Neuroscience*, 7, 151.
- Megemont, M., McBurney-Lin, J. & Yang, H. (2022). Pupil diameter is not an accurate real-time readout of locus coeruleus activity. *Elife*, 11, e70510.
- Miller, M. D. & MacKay, D. G. (1994). Repetition deafness: Repeated words in computer-compressed speech are difficult to encode and recall. *Psychological Science*, 5(1), 47–51.
- Miyazawa, S. & Iwasaki, S. (2010). Do happy faces capture attention? the happiness superiority effect in attentional blink. *Emotion*, 10(5), 712.
- Mostofsky, S. H. & Simmonds, D. J. (2008). Response inhibition and response selection: Two sides of the same coin. *Journal of Cognitive Neuroscience*, 20(5), 751–761.
- Müllensiefen, D., Gingras, B., Musil, J. & Stewart, L. (2014a). Measuring the facets of musicality: The Goldsmiths Musical Sophistication Index (Gold-MSI). *Personality and Individual Differences*, 60, S35.
- Müllensiefen, D., Gingras, B., Musil, J. & Stewart, L. (2014b). The musicality of non-musicians: an index for assessing musical sophistication in the general population. *PLoS One*, 9(2), e89642.
- Murray, M. M., Camen, C., Andino, S. L. G., Bovet, P. & Clarke, S. (2006). Rapid brain discrimination of sounds of objects. *Journal of Neuroscience*, 26(4), 1293–1302.
- Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasley, B. & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience*, 15(7), 1040–1046.
- Nesse, R. M. (2019). *Good reasons for bad feelings: Insights from the frontier of evolutionary psychiatry*. UK: Allen Lane.
- Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*, 5(2), 241–301.
- Nieuwenhuis, S., Gilzenrat, M. S., Holmes, B. D. & Cohen, J. D. (2005). The role of the locus coeruleus in mediating the attentional blink: a neurocomputational theory. *Journal of Experimental Psychology: General*, 134(3), 291.
- Nieuwenhuis, S., van Nieuwpoort, I. C., Veltman, D. J. & Drent, M. L. (2007). Effects of the noradrenergic agonist clonidine on temporal and spatial attention. *Psychopharmacology*, 193(2), 261–269.

Bibliography

- Norman-Haignere, S. V., Feather, J., Boebinger, D., Brunner, P., Ritaccio, A., McDermott, J. H., Schalk, G. & Kanwisher, N. (2022). A neural population selective for song in human auditory cortex. *Current Biology*, *32*(7), 1470–1484.
- Olivers, C. N. & Nieuwenhuis, S. (2005). The beneficial effect of concurrent task-irrelevant mental activity on temporal attention. *Psychological Science*, *16*(4), 265–269.
- Olivers, C. N. & Nieuwenhuis, S. (2006). The beneficial effects of additional task load, positive affect, and instruction on the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(2), 364.
- Ollen, J. E. (2006). *A criterion-related validity test of selected indicators of musical sophistication using expert ratings*. PhD thesis, The Ohio State University.
- Palmer, C. & Drake, C. (1997). Monitoring and planning capacities in the acquisition of music performance skills. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *51*(4), 369–384.
- 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.
- Pantev, C., Roberts, L. E., Schulz, M., Engelien, A. & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport*, *12*(1), 169–174.
- Parbery-Clark, A., Skoe, E., Lam, C. & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing*, *30*(6), 653–661.
- Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, *116*(2), 220–224.
- Patching, G. (2018). Bayes factors: A ‘re-revolution’ in psychology. In *Bayes@Lund 2018*.
- Pecchinenda, A., Monachesi, B. & Laeng, B. (2020). Fearful expressions of rapidly presented hybrid-faces modulate the lag 1 sparing in the attentional blink. *Acta Psychologica*, *209*, 103124.
- Pernet, C. R., McAleer, P., Latinus, M., Gorgolewski, K. J., Charest, I., Bestelmeyer, P. E., Watson, R. H., Fleming, D., Crabbe, F., Valdes-Sosa, M. et al. (2015). The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. *Neuroimage*, *119*, 164–174.
- Perrodin, C., Kayser, C., Logothetis, N. K. & Petkov, C. I. (2011). Voice cells in the primate temporal lobe. *Current Biology*, *21*(16), 1408–1415.

- Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M. & Logothetis, N. K. (2008). A voice region in the monkey brain. *Nature Neuroscience*, *11*(3), 367–374.
- Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(5), 976–986.
- Potter, M. C., Chun, M. M., Banks, B. S. & Muckenhoupt, M. (1998). Two attentional deficits in serial target search: The visual attentional blink and an amodal task-switch deficit. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(4), 979–992.
- Potter, M. C. & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology*, *81*(1), 10–15.
- Preuschoff, K., Hart, B. M. & Einhauser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Frontiers in Neuroscience*, *5*, 115.
- Privitera, C. M., Renninger, L. W., Carney, T., Klein, S. & Aguilar, M. (2010). Pupil dilation during visual target detection. *Journal of Vision*, *10*(10), 3.
- Rajkowski, J., Kubiak, P. & Aston-Jones, G. (1993). Correlations between locus coeruleus (lc) neural activity, pupil diameter and behavior in monkey support a role of lc in attention. *Society of Neuroscience Abstracts* (p. 974).
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510–532.
- Raymond, J. E., Shapiro, K. L. & Arnell, K. M. (1992). Temporary suppression of visual processing in an rsvp task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 849–860.
- Reimer, J., Froudarakis, E., Cadwell, C. R., Yatsenko, D., Denfield, G. H. & Tolias, A. S. (2014). Pupil fluctuations track fast switching of cortical states during quiet wakefulness. *Neuron*, *84*(2), 355–362.
- Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A. & Tolias, A. S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature Communications*, *7*(1), 1–7.
- Rendall, D., Rodman, P. S. & Emond, R. E. (1996). Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Animal Behaviour*, *51*(5), 1007–1015.
- Rensink, R. A., O'Regan, J. K. & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*(5), 368–373.

Bibliography

- Rodrigues, A. C., Loureiro, M. A. & Caramelli, P. (2013). Long-term musical training may improve different forms of visual attention ability. *Brain and Cognition*, *82*(3), 229–235.
- Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, *21*(2), 301–308.
- Samuels, E. R. & Szabadi, E. (2008). Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part i: principles of functional organisation. *Current Neuropharmacology*, *6*(3), 235–253.
- Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews Neuroscience*, *10*(3), 211–223.
- Schellenberg, E. G. (2005). Music and cognitive abilities. *Current Directions in Psychological Science: A Journal of the American Psychological Society*, *14*(6), 317–320.
- Schellenberg, E. G. & Peretz, I. (2007). Music, language and cognition: Unresolved issues. *Trends in Cognitive Sciences*, *12*(2), 45–46.
- Schulze, K., Koelsch, S. & Williamson, V. (2018). Auditory working memory. In *Springer handbook of systematic musicology* (pp. 461–472). Springer.
- Shapiro, K. (2001). Temporal methods for studying attention: How did we get here and where are we going? In *The Limits of Attention: Temporal Constraints in Human Information Processing*. Oxford: Oxford University Press.
- Shapiro, K. L., Caldwell, J. & Sorensen, R. E. (1997a). Personal names and the attentional blink: A visual "cocktail party" effect. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(2), 504–514.
- Shapiro, K. L., Raymond, J. & Arnell, K. (1997b). The attentional blink. *Trends in Cognitive Sciences*, *1*(8), 291–296.
- Shelton, J. T., Elliott, E. M., Hill, B., Calamia, M. R. & Gouvier, W. D. (2009). A comparison of laboratory and clinical working memory tests and their prediction of fluid intelligence. *Intelligence*, *37*(3), 283–293.
- Shen, D. & Alain, C. (2010). Neuroelectric correlates of auditory attentional blink. *Psychophysiology*, *47*(1), 184–191.
- Shen, D. & Alain, C. (2011). Temporal attention facilitates short-term consolidation during a rapid serial auditory presentation task. *Experimental Brain Research*, *215*(3), 285–292.
- Shen, D. & Alain, C. (2012). Implicit temporal expectation attenuates auditory attentional blink. *PLoS One*, *7*(4), e36031.

- Shen, D. & Mondor, T. A. (2006). Effect of distractor sounds on the auditory attentional blink. *Perception & Psychophysics*, *68*(2), 228–243.
- Simons, D. J. & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentive blindness for dynamic events. *Perception*, *28*(9), 1059–1074.
- Slagter, H. A., Lutz, A., Greischar, L. L., Francis, A. D., Nieuwenhuis, S., Davis, J. M. & Davidson, R. J. (2007). Mental training affects distribution of limited brain resources. *PLoS Biology*, *5*(6), e138.
- Slawinski, E. B. & Goddard, K. M. (2001). Age-related changes in perception of tones within a stream of auditory stimuli: Auditory attentional blink. *Canadian Acoustics*, *29*(1), 3–12.
- Slevc, L. R., Davey, N. S., Buschkuhl, M. & Jaeggi, S. M. (2016). Tuning the mind: Exploring the connections between musical ability and executive functions. *Cognition*, *152*, 199–211.
- Smith, J. (2008). *Vocal Tracks: Performance and Sound Media*. Berkeley: University of California Press.
- Soto-Faraco, S. & Spence, C. (2002). Modality-specific auditory and visual temporal processing deficits. *The Quarterly Journal of Experimental Psychology*, *55*(1), 23–40.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, *30*, 276–315.
- 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.
- Suied, C., Agus, T. R., Thorpe, S. J., Mesgarani, N. & Pressnitzer, D. (2014). Auditory gist: Recognition of very short sounds from timbre cues. *The Journal of the Acoustical Society of America*, *135*(3), 1380–1391.
- Suied, C., Agus, T. R., Thorpe, S. J. & Pressnitzer, D. (2013). Processing of short auditory stimuli: The rapid audio sequential presentation paradigm (RASP). In *Basic Aspects of Hearing* (pp. 443–451). Springer.
- Szabadi, E. (2012). Modulation of physiological reflexes by pain: role of the locus coeruleus. *Frontiers in Integrative Neuroscience*, *6*, 94.
- Taatgen, N. A., Juvina, I., Schipper, M., Borst, J. P. & Martens, S. (2009). Too much control can hurt: A threaded cognition model of the attentional blink. *Cognitive Psychology*, *59*(1), 1–29.
- Tabachnick, B. G., Fidell, L. S. & Ullman, J. B. (2007). *Using multivariate statistics*, Volume 5. Boston: Pearson/Allyn and Bacon.

Bibliography

- Tang, M. F., Badcock, D. R. & Visser, T. A. (2014). Training and the attentional blink: Limits overcome or expectations raised? *Psychonomic Bulletin & Review*, *21*(2), 406–411.
- Taylor, T. L. (2018). Remember to blink: Reduced attentional blink following instructions to forget. *Attention, Perception, & Psychophysics*, *80*(6), 1489–1503.
- Tervaniemi, M. (2009). Musicians—same or different? *Annals of the New York Academy of Sciences*, *1169*(1), 151–156.
- Tremblay, S., Vachon, F. & Jones, D. M. (2005). Attentional and perceptual sources of the auditory attentional blink. *Perception & Psychophysics*, *67*(2), 195–208.
- Troche, S. J. & Rammsayer, T. H. (2013). Attentional blink and impulsiveness: Evidence for higher functional impulsivity in non-blinkers compared to blinkers. *Cognitive Processing*, *14*(3), 273–281.
- Vachon, F. & Tremblay, S. (2005). Auditory attentional blink: Masking the second target is necessary, delayed masking is sufficient. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *59*(4), 279–286.
- Vitevitch, M. S. (2003). Change deafness: The inability to detect changes between two voices. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 333–342.
- Vogel, E. K. & Luck, S. J. (2002). Delayed working memory consolidation during the attentional blink. *Psychonomic Bulletin & Review*, *9*(4), 739–743.
- Vouloumanos, A., Hauser, M. D., Werker, J. F. & Martin, A. (2010). The tuning of human neonates' preference for speech. *Child Development*, *81*(2), 517–527.
- Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779–804.
- Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C. & Vuust, P. (2010). The Musical Ear Test, a new reliable test for measuring musical competence. *Learning and Individual Differences*, *20*(3), 188–196.
- Webb, A. R., Heller, H. T., Benson, C. B. & Lahav, A. (2015). Mother's voice and heartbeat sounds elicit auditory plasticity in the human brain before full gestation. *Proceedings of the National Academy of Sciences*, *112*(10), 3152–3157.
- Wechsler, D. (2003). *Manual for the Wechsler Adult Intelligence Scale—Third Edition, Norwegian version*. Stockholm: Pearson.

- Weiss, M. W., Trehub, S. E. & Schellenberg, E. G. (2012). Something in the way she sings: Enhanced memory for vocal melodies. *Psychological Science*, *23*(10), 1074–1078.
- Welford, A. T. (1952). The psychological refractory period and the timing of high-speed performance - a review and a theory. *British Journal of Psychology*, *43*(1), 2–19.
- Whitehead, J. C. & Armony, J. L. (2018). Singing in the brain: Neural representation of music and voice as revealed by fMRI. *Human Brain Mapping*, *39*(12), 4913–4924.
- Wierda, S. M., van Rijn, H., Taatgen, N. A. & Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences*, *109*(22), 8456–8460.
- Wierda, S. M., Van Rijn, H., Taatgen, N. A. & Martens, S. (2010). Distracting the mind improves performance: An ERP study. *PloS One*, *5*(11), e15024.
- Willems, C., Damsma, A., Wierda, S. M., Taatgen, N. & Martens, S. (2015a). Training-induced changes in the dynamics of attention as reflected in pupil dilation. *Journal of Cognitive Neuroscience*, *27*(6), 1161–1171.
- Willems, C., Herdizin, J. & Martens, S. (2015b). Individual differences in temporal selective attention as reflected in pupil dilation. *PloS One*, *10*(12), e0145056.
- Willems, C. & Martens, S. (2016). Time to see the bigger picture: Individual differences in the attentional blink. *Psychonomic Bulletin & Review*, *23*(5), 1289–1299.
- Wong, N. H. & Chang, D. H. (2018). Attentional advantages in video-game experts are not related to perceptual tendencies. *Scientific Reports*, *8*(1), 5528–9.
- Young, A. W., Frühholz, S. & Schweinberger, S. R. (2020). Face and voice perception: Understanding commonalities and differences. *Trends in Cognitive Sciences*, *24*(5), 398–410.
- Yovel, G. & Belin, P. (2013). A unified coding strategy for processing faces and voices. *Trends in Cognitive Sciences*, *17*(6), 263–271.
- Zatorre, R. J., Chen, J. L. & Penhune, V. B. (2007). When the brain plays music: Auditory–motor interactions in music perception and production. *Nature Reviews Neuroscience*, *8*(7), 547–558.
- Zylberberg, A., Oliva, M. & Sigman, M. (2012). Pupil dilation: A fingerprint of temporal selection during the “attentional blink”. *Frontiers in Psychology*, *3*, 316.

Papers

Paper I

No evidence for an auditory attentional blink for voices regardless of musical expertise

Merve Akça, Bruno Laeng, & Rolf Inge Godøy

Published in *Frontiers in Psychology*, January 2020, volume 10, pp. 2935. DOI: 10.3389/fpsyg.2019.02935.



No Evidence for an Auditory Attentional Blink for Voices Regardless of Musical Expertise

Merve Akça^{1,2*}, Bruno Laeng^{1,3} and Rolf Inge Godoy^{1,2}

¹ RITMO Center for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo, Oslo, Norway, ² Department of Musicology, University of Oslo, Oslo, Norway, ³ Department of Psychology, University of Oslo, Oslo, Norway

Background: Attending to goal-relevant information can leave us metaphorically “blind” or “deaf” to the next relevant information while searching among distracters. This temporal cost lasting for about a half a second on the human selective attention has been long explored using the attentional blink paradigm. Although there is evidence that certain visual stimuli relating to one’s area of expertise can be less susceptible to attentional blink effects, it remains unexplored whether the dynamics of temporal selective attention vary with expertise and objects types in the auditory modality.

Methods: Using the auditory version of the attentional blink paradigm, the present study investigates whether certain auditory objects relating to musical and perceptual expertise could have an impact on the transient costs of selective attention. In this study, expert cellists and novice participants were asked to first identify a target sound, and then to detect instrumental timbres of cello or organ, or human voice as a second target in a rapid auditory stream.

Results: The results showed moderate evidence against the attentional blink effect for voices independent of participants’ musical expertise. Experts outperformed novices in their overall accuracy levels of target identification and detection, reflecting a clear benefit of musical expertise. Importantly, the musicianship advantage disappeared when the human voices served as the second target in the stream.

Discussion: The results are discussed in terms of stimulus salience, the advantage of voice processing, as well as perceptual and musical expertise in relation to attention and working memory performances.

Keywords: attentional blink, temporal selective attention, perceptual expertise, musical expertise, auditory attention

1. INTRODUCTION

Attentional blink (AB; Raymond et al., 1992) refers to the phenomenon that when two targets are presented in close temporal proximity, report of the second target (T2) is often impaired after correctly identifying the first target (T1). The AB is known to be an effective tool to study the time-course of selective attention and has been studied extensively in the visual domain using a wide range of material. However, to date, relatively few studies have investigated its auditory counterpart. Those investigations have shown evidence for the auditory AB effect (but see Potter et al., 1998 for

OPEN ACCESS

Edited by:

Claude Alain,
Rotman Research Institute (RRI),
Canada

Reviewed by:

Dawei Shen,
Rotman Research Institute (RRI),
Canada
Kimmo Alho,
University of Helsinki, Finland

*Correspondence:

Merve Akça
merve.akca@imv.uio.no

Specialty section:

This article was submitted to
Auditory Cognitive Neuroscience,
a section of the journal
Frontiers in Psychology

Received: 03 September 2019

Accepted: 11 December 2019

Published: 10 January 2020

Citation:

Akça M, Laeng B and Godoy RI
(2020) No Evidence for an Auditory
Attentional Blink for Voices Regardless
of Musical Expertise.
Front. Psychol. 10:2935.
doi: 10.3389/fpsyg.2019.02935

conflicting evidence) using simple, non-musical tones (e.g., Shen and Mondor, 2006; Shen and Alain, 2012), spoken letters and digits (e.g., Arnell and Larson, 2002; Arnell and Jenkins, 2004; Martens et al., 2015), and spoken syllables (e.g., Duncan et al., 1997; Tremblay et al., 2005) as auditory stimuli. The aim of the present study is to investigate temporal dynamics of selective auditory attention and whether these dynamics can be modulated by an individual's expertise and the objects of this expertise. We focused in particular on common perceptual expertise with voices and musical expertise. Here the perceptual expertise with voices refers to the common human experience that is associated with higher level of perceptual capacity for voices, shaped by extensive exposure to human voices for years on a daily basis, while the musical expertise refers to advanced levels of music performance experience combined with extensive training in music. Employing the auditory version of the attentional blink paradigm, we explored the temporal costs of selective attention among expert cellist and novice participants for different auditory objects through systematically manipulating the type of the second target (human voice, cello, organ) and the interval between the first and second target (i.e., lag) presented in a rapid auditory stream.

Only recently has it been demonstrated that AB is not as universal of a cognitive limitation as once was thought, but that it greatly differs between individuals, or groups of individuals, and these differences in the AB frequency are dependent on various factors, such as stimulus category, duration, and modality (Willems and Martens, 2016). Previous studies also illustrated that certain visual stimuli, such as faces (Awh et al., 2004; Landau and Bentin, 2008), which constitute an expertise object for almost all of us, and other expertise objects (e.g., cars for car experts, Blacker and Curby, 2016) can lower one's susceptibility to the AB effects. Less well-understood is whether this kind of expertise-related alleviation in the AB effects can be observed in the auditory domain, since the few existing studies regarding auditory AB did not explore this factor.

In this study human voices were used as it is believed that they are, as the human faces in the visual modality, objects of expertise for most of us. Ample evidence exists supporting the neural and cognitive similarities in the perception of human faces and voices, suggesting a unifying coding mechanism (for a review, see Yovel and Belin, 2013). It can be argued that due to our extensive experience with (and the evolutionary importance of) human voices, all humans with normal hearing abilities can be considered voice experts. To contrast the possible AB effects (or lack thereof) for human voices, instrumental tones were used, which were cello and organ tones. Interestingly, cello tones have the strongest acoustic similarity to human voices among instruments, as reported by listeners in Askenfelt (1991)'s study. Later neuroimaging studies also support that cello tones share some similarities with the human voice. For example (Levy et al., 2003), showed that string instruments elicited a larger positivity than brass and woodwind instruments, with cello evoking the largest positivity among the string instruments. The authors suggested that these findings may reflect the perceptual similarity between string instruments and human voices. Furthermore, a recent fMRI study also indicated a direct overlap on the brain

networks during cello playing and during singing, as well as that playing cello may directly engage the vocal areas of the brain especially for the experienced cellists whose musical training started before the age of 7 (Segado et al., 2018). It is therefore noteworthy to explore the AB effects for cello tones not only as an expertise-object for the cellists but also as a possible expertise object for the people who has no musical training due to the similarities shared with the human voices. Organ tones were chosen as a control for both cello tones and human voices, as a contrasting timbre with an assumed common familiarity with organ tones for individuals from most Western cultures without a professional expertise as an organist.

Defined as the ability of accurate and rapid identification of individual sound sources within a set of homogenous stimuli (Chartrand et al., 2008), auditory recognition expertise has been studied mostly with musicians, and musicians in this sense are often considered as "auditory experts." There exists evidence that musicians' auditory processing abilities differ from that of non-musicians at a behavioral level. For example, tasks involving pitch discrimination, processing of temporal information, processing of instrumental timbres, and discrimination of instrumental and voice timbres (Chartrand and Belin, 2006), all have been observed to be performed better among musicians than non-musicians. Furthermore, according to some neuroimaging studies, musician and non-musician brains seem to have differences in terms of volume, density, connectivity, morphology, and functional activity across various brain regions and structures (e.g., Gaser and Schlaug, 2003; Hyde et al., 2009). Although it is often difficult to disentangle the influences of external environmental factors from the innate biological factors, some longitudinal studies seem to suggest that the differences in musicians' and non-musicians' brain reflect their learning experiences (e.g., Hyde et al., 2009). However, it is important to note that the present cross-sectional study does not aim at differentiating training-related and genetic factors in becoming an expert musician, or asserts the observed musical expertise benefits to have a genetic or environmental basis.

The question of whether transfer of the skill, that is, the ability of specific experience to impact seemingly unrelated processes, is also explored in several studies (see Kraus and Chandrasekaran, 2010 for a review). In the context of musical training, both near transfer (i.e., benefits in highly similar contexts/domains despite the lack of training, such as better perception of piano tones in violinists) and far transfer (i.e., benefits to activities outside of the trained domain, such as language processing) effects has been documented (Moreno and Bidelman, 2014). So musical training may have benefits that are not simply limited within the scope of music field but often extends to influencing high-level cognitive functions, such as cognitive control, attention and working memory (WM). For example, a recent study pointed out that the association between music training and executive functioning was strongest for the executive functioning of WM in both visual and auditory modalities (Slevc et al., 2016). This, by itself, however does not ascertain far-transfer of musical training skills, as causality cannot be inferred from correlational studies.

Selective attention and working memory are known to share a strong link. Considering the AB literature, interaction of WM

and the AB has been demonstrated behaviorally (e.g., Akyürek et al., 2007) and with brain imaging techniques (e.g., Johnston et al., 2012). Moreover, individuals with higher levels of WM functioning and broad attentional focus seem to perform better in the AB paradigm compared to those with lower WM and with narrow attentional focus (for a review see Willems and Martens, 2016). There are some documentations in the literature which may indicate better deployment of attention in time among musicians than non-musicians. For example, using tones (auditory AB task stimuli) and lines (visual AB task stimuli), Slawinski et al. (2002) reported attenuated AB in both visual and auditory modalities in musicians compared to non-musicians. Using letters and digits (presented in auditory and visual modalities), Martens et al. (2015) observed an attenuation and delay of the AB only in the auditory modality in musicians and suggested that music training have a modality-specific beneficial effect on selective attention. Despite these documentations of attenuated auditory AB in musicians, the question of whether this benefit on auditory attention can be altered using other kinds of auditory stimuli remains unanswered.

The four main research questions posed in the present experiment are as follows:

1. Are human voices less susceptible to the auditory AB effects than instrumental tones?
2. If so, does this effect extend to instruments sharing perceptual similarity to human voice (i.e., cello tones)?
3. Do expert musicians show an overall attenuation of the auditory AB as compared with novices due to their extensive auditory training with tones?
4. Do expert musicians show less auditory AB to the musical timbres associated with their principal instrument compared to timbres associated with the instruments that they have not been trained on?

Additionally, we explored whether musical sophistication (as measured by Gold-MSI inventory) and WM span (as measured with Letter-Number Sequencing task) are related to individual or group level differences in the auditory AB effect.

2. MATERIALS AND METHODS

2.1. Participants

Thirty-eight volunteers (19 expert cellists and 19 novice participants) were recruited for the present study. Novice participants were defined as participants with no previous cello training and with very little or no musical training, while the expert cellists were defined as professional or advanced cello players with extensive musical training. Data from one novice participant were removed due to having a T1 identification accuracy rate below the chance level. Data from one cellist with a musical sophistication score and years of cello experience $>2SD$ below the group mean was also not included in the analysis. The final sample consisted of 18 expert cellists (13 female, age range = 21–36 years, mean age = 28.89, $SD = 5.06$) and 18 novices (13 female, age range = 21–39 years, mean age = 28.17, $SD = 5.86$). The groups did not differ significantly in terms of their age [$t_{(34)} = -0.396$, $p = 0.69$] and no statistically significant

relationship was observed between participant group and gender ($p = 1.00$, odds ratio = 0.00, CI%: -1.459 to 1.459). The average number of years of cello experience was 19.94 years ($SD = 5.25$, ranging from 9 to 28 years). The majority of the expert cellists was recruited from the Norwegian Academy of Music and the novices were recruited from the University of Oslo. The general exclusion criteria were having a history of hearing, speech or neurological disorders. All participants received a gift card worth of 200 NOK as compensation for their time.

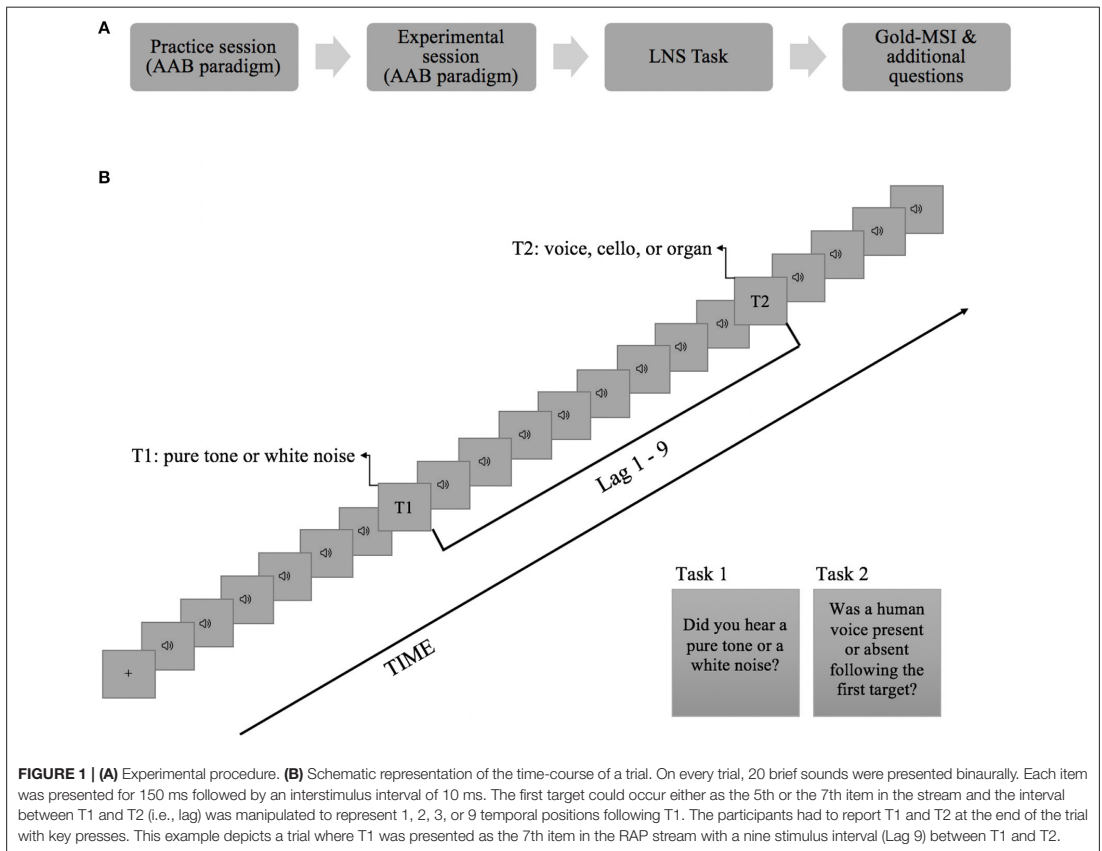
2.2. Auditory Stimuli

Auditory stimuli consisted of 16 human voice excerpts, 16 cello tones, 16 organ tones, as well as 16 pure sine tones at various frequencies equally spaced on a logarithmic scale, white noise, and 48 different environmental sounds (e.g., broom, doorbell, motorbike). Instrumental sounds, voices, and sine tones were periodic and with harmonic spectra, while the environmental sounds and white noise were mostly non-periodic and did not have harmonic spectra. Pure tones and white noise were generated using Audacity 2.2.1 sound editing software (<https://audacityteam.org/>). Cello and organ (i.e., baroque plenum) instrument sounds were sampled from the McGill University Master Samples DVD set (Opolko and Wapnick, 2006), and the rest of the stimuli were sampled from freesounds.org. We have chosen 150 ms excerpts from the quasi-stationary portions of the sound (thus, avoided the initial attack transients). The excerpts were taken for cello, organ, and voice stimuli on D#2, E2, F2, G2, G#2, C3, C#3, D3, D#3, E3, F3, F#3, G3, G#3, A#3, and B3 pitches. Human voices included digital recordings of vowels sung by a male voice. Stimuli were normalized using the peak normalization method and matched by duration (150 ms with 2 ms linear amplitude ramps to eliminate the perceptual effects of onset and offset clicks). In the experimental paradigm, pure tones and white noise served as T1, human voice, cello, organ tones served as T2, and environmental sounds served as distracters.

2.3. Design and Procedure

The experiment took place at the cognitive laboratory in the Department of Psychology in the University of Oslo. It started with a practice session which then was followed by an experimental session of the AB task. E-Prime 3.0 software (Psychology Software Tools, Pittsburgh, PA, USA) was used for the presentation of the stimuli and for response collection in the AB task. After completion of the AB task, the Letter-Number Sequencing task and the Musical Sophistication Index were administered (see **Figure 1A** for an illustration of the experimental procedure). The practice session consisted of 16 trials representing all T2 categories and the participants were given accuracy feedback at the end of every trial. In the experimental session, the participants went through a total of 192 trials with three alternating blocks of all T2 types. The order of the blocks was counterbalanced across participants. No feedback was provided to the participants during the experimental session.

Figure 1B depicts an example trial from the AB task. Each trial began with a fixation cross (+) appearing at the center of the screen for 250 ms. Then, a rapid auditory presentation



stream (RAP) consisting of 20 items (distracters and targets) was presented binaurally via Beyerdynamic DT770 Pro circumaural headphones. Each item in the stream lasted for 150 ms followed by an inter-stimulus interval of 10 ms, yielding a presentation rate of 6.25 Hz. T1 was randomly selected from the pool of pure tones (on 50% of the trials) and noise bursts (50% of the trials), and was placed in either the 5th or 7th serial position in the auditory stream. While T1 was presented on every trial, T2 was randomly presented only on 50% of the trials and its pitch and lag from T1 were randomly varied. T2 could appear at Lag1, Lag2, Lag3, or Lag9 following the T1 (corresponding to 160, 320, 480, or 1440 ms of stimulus onset asynchrony between the targets, respectively). When present, in 32 trials T2 was a cello tone, in 32 trials it was a human voice (i.e., sung tone), and in 32 trials it was an organ tone.

At the end of each trial, participants were first asked to identify the first target with a two-alternative forced choice method (pure tone or noise) and then to detect the second target item (e.g., “Was a human voice present or absent following the first target?”). Responses were given by pressing the corresponding

keys on the computer keyboard. Participants were instructed to take their time in making their responses.

2.4. Musical Sophistication Index

Self-report inventory of the Goldsmiths Musical Sophistication Index (Gold-MSI version 1.0; Müllensiefen et al., 2014) was used to quantify participants' self-reported musical sophistication. The Gold-MSI inventory comprises five sub-scales (active engagement, perceptual abilities, musical training, singing abilities, and emotion) and one general factor (i.e., musical sophistication). It contains 38 items rated on a seven-point scale (and an additional question of the instrument played best). The Gold-MSI inventory is suited for measuring musical engagement and behavior in the general population (i.e., not restricted to the musically trained group).

All scales of the Gold-MSI inventory has been shown to have a good internal consistency (Cronbach's alpha = 0.926 for the general musical sophistication factor, 0.872 for active engagement, 0.873 for perceptual abilities, 0.903 for musical training, 0.870 for singing abilities, 0.791 for

emotions; Müllensiefen et al., 2014) a very high re-test reliability (ranging between 0.857 and 0.972; Müllensiefen et al., 2014). The inventory has been demonstrated to have convergent validity with the musical aptitude subscale of the Musical Engagement Questionnaire (MEQ; Werner et al., 2006) and discriminant validity with the less related MEQ subscales like affective reactions ($r = 0.449$, $p < 0.01$ and $r = 0.182$, $p < 0.05$, respectively).

2.5. Letter-Number Sequencing Task

Letter-number sequencing (LNS) is a supplemental subtest of the Wechsler Adult Intelligence Test (Wechsler, 2008). The LNS is a complex WM span task which involves hearing a sequence of numbers and letters, and then reporting back the numbers in ascending order and the letters in alphabetic order. The task requires not only relying on attention and auditory memory but also manipulating auditory information. Together with the digit span, LNS was found the most highly related psychometric tests to laboratory WM measures (Shelton et al., 2009). Fluid intelligence and cognitive flexibility have also been argued to be involved by this task (Pezzuti and Rossetti, 2017). The individual differences in the temporal costs of selective visual attention has been previously linked to WM capacity and fluid intelligence was associated with higher target accuracy (Colzato et al., 2007; but see Martens and Johnson, 2009 for contrasting evidence) using other tasks (e.g., Operation Span task, Symmetry, and Reading Span tasks as measures of WM capacity, and Raven's Progressive Matrices to assess fluid intelligence), but not with LNS complex span task. It is therefore interesting to explore whether the performance in LNS could be linked to individual or group level differences in the auditory AB. The auditory nature of the task makes it additionally suitable for this study.

2.6. Data Analyses

Performance on the T1 identification task was assessed using a repeated-measures analysis of variance (ANOVA) with a between-subjects variable of Group (novice, expert) and the within subject variables of T2 Type (voice, cello, organ) on T1 accuracy data. To explore the AB effects, T2|T1 accuracy (i.e., the accuracy of T2 detection when T1 was correctly reported) data were analyzed using repeated-measures ANOVAs with a between-subjects variable of Group and the within subject variables of T2 Type and Lag. Split-up ANOVAs were performed when interactions were observed. Greenhouse-Geisser and Huynh-Feldt corrections were applied when the sphericity assumption was not met. The significance threshold was set to $p < 0.05$ for all tests. In addition, Bayesian statistical analyses were conducted to determine the evidence proportion in favor of the null (i.e., the absence of an effect, H_0) and alternative (H_1) hypotheses. Bayesian statistics were computed using JASP (JASP Team, 2018; jasp-stats.org). For all the reported Bayesian ANOVAs, the default prior of 0.5 was used for the r-scaled fixed effects as no other information was available to update this prior.

Bayes factors (BF10) reflect the likelihood of the data to arise from alternative model (H_1) in comparison to the null model (H_0). In some of the analyses with a priori contrasts, the alternative hypothesis was formulated as a directional hypothesis

and the Bayes factor reported in these analyses as BF-0. Bayes Inclusion Factor across matched-models (also known as Baws factor; Mathôt, 2017) was used when reporting and interpreting the results obtained from the Bayesian ANOVA with the multifactorial models design. Bayes Inclusion Factor (BF_{Inclusion}) across matched models reflects the evidence for all models that includes a certain effect to equivalent models stripped of that effect. The present interpretation of the Bayes factors for evidence is based on the recommendations by Dienes (2014). Especially, when the BF is comprised between 0 and 0.33, one can confidently conclude in favor of the null model (i.e., no difference). When the BF is between 0.33 and 3 the evidence is inconclusive and no specific statement can be made about which model is supported by the evidence. When BF is 3 or above (to infinity), we have conclusive evidence for a difference. Importantly, the Bayesian approach allows to specify the relative probability that either the alternative or null hypotheses are true. We also consulted Lee and Wagenmakers (2014)'s adjustment of the Jeffreys (1961) original labeling of the evidence categories, which can be found in **Supplementary Material**.

3. RESULTS

3.1. T1 Accuracy

The overall performance on T1 identification was high across all three T2 conditions (90, 89, and 88% in voice, cello, and organ conditions, respectively). A repeated measures ANOVA showed a significant main effect of Group, $F_{(1, 34)} = 12.03$, $p = 0.001$, $\eta_p^2 = 0.261$. *Post-hoc* contrasts revealed that expert cellists were better at correctly identifying T1 than the novices (Mean difference = -0.136 , $SE = 0.039$). The main effect of T2 Type, $F_{(2, 68)} = 0.751$, $p = 0.476$, $\eta_p^2 = 0.022$ and the interaction of T2 Type and Group was not significant, $F_{(2, 68)} = 1.074$, $p = 0.347$, $\eta_p^2 = 0.031$. As non-significant results cannot be interpreted as a conclusive evidence for the absence of an effect (Dienes, 2014), Bayesian analyses were necessary. As shown in **Table 1**, Bayesian testing on T1 accuracy scores indicated a moderate evidence supporting that T1 accuracy is 6.17 times more likely to not be affected by T2 Type (BF_{Inclusion} = 0.162). This speaks against the differential processing of T1 in the context of voice, cello, and organ detection. Similarly, the null model received moderate support in explaining the data as compared to the interaction of the two factors, BF_{Inclusion} = 0.307. Finally, the estimated Bayes factor revealed a strong evidence supporting that the differences in T1 accuracy was about 20 times more likely to occur under the model with participant group than the null model, BF_{Inclusion} = 20.365.

3.2. T2|T1 Accuracy

T2|T1 performance analysis (based on trials in which T1 was correctly identified) as a function of lag is the attentional blink effects critical to this study's aim. Results of three-way ANOVAs on T2|T1 performance across all lag conditions are summarized in **Table 2**. In the following analyses, we focus on the T2|T1 accuracy at Lag 3 (inside the typical AB time window) and at Lag 9 (outside the AB time-window). Lag 3 was selected due to yielding the lowest performance among the lags within the typical

TABLE 1 | Bayesian analysis of variance across matched models with T2 Type (voice, cello, organ) and Group (novice, cellist) on T1 accuracy scores.

Effects	P (incl)	P (incl data)	BF incl
Analysis of effects			
T2 type	0.400	0.131	0.162
Group	0.400	0.916	20.365
T2 type × group	0.200	0.038	0.307

The analysis compares models that contain the effect to equivalent models stripped of the effect. Higher-order interactions are excluded.

AB period (< 500 ms). If we were to observe an attentional blink, it would be where the T2 deficit is largest within the typical AB period, and that it would recover outside of this time window.

A three-way ANOVA with T2 Type (voice, cello, organ), Lag (3, 9), and Group (novice, cellist) showed a significant main effect of T2 Type [$F_{(2, 68)} = 9.68, p < 0.001, \eta_p^2 = 0.222$], Group [$F_{(1, 34)} = 17.63, p < 0.001, \eta_p^2 = 0.341$], and the interaction between T2 Type and Group [$F_{(2, 68)} = 9.661, p < 0.001, \eta_p^2 = 0.221$]. *Post-hoc* comparison of T2 Type using Bonferroni procedure revealed that, on the condition that T1 being correctly reported, the voice targets were more accurately reported than the cello targets ($p = 0.006$). T2|T1 accuracy on the voice targets were also higher than the organ targets but this contrast was at the margin of statistical significance ($p = 0.71$). *Post-hoc* comparison of Group revealed that the cellists had significantly higher overall T2|T1 accuracy than the novices ($p < 0.01$). The main effect of Lag ($p = 0.22, \eta_p^2 = 0.044$) and the interaction effects with Lag (Lag × Group, $p = 0.12, \eta_p^2 = 0.068$; T2 Type × Lag, $p = 0.60, \eta_p^2 = 0.013$; T2 Type × Lag × Group, $p = 0.12, \eta_p^2 = 0.061$) did not reach statistical significance. A parallel Bayesian ANOVA showed anecdotal evidence against the Lag-only model (BFInclusion = 0.284). There was inconclusive evidence against the model with the interaction Lag and Group (BFInclusion = 0.637), while the model with the T2 Type × Lag interaction reflected substantial evidence against the interaction, favoring the null hypothesis roughly 9 times more (BFInclusion = 0.110). There was inconclusive evidence against the model explaining the data with the three-way interaction of T2 Type × Lag × Group (BFInclusion = 0.370). The estimated Bayes factor revealed extreme evidence in favor of the models including the T2 Type-only (BFInclusion = 1236.378), Group-only (BFInclusion = 128.94), and the interaction of T2 Type and Group (BFInclusion = 4807.098) than the null model. The model that received the most support was that of the interaction of T2 Type and Group, which was around 4807 times more likely to explain the data than the null model did.

Figure 2 illustrates the T2|T1 accuracy percentages of expert cellists and novices under T2 conditions across the two critical Lag conditions (Lag 3 and Lag 9). As visible in the figure, there appears to be an overall benefit of musical expertise across T2 conditions. It also appears that the human voices were less susceptible to the auditory attentional blink regardless of musical expertise. To further explore the two-way interaction of T2 Type and Group, split-up ANOVAs were conducted.

3.2.1. T2 Type Differences on T2|T1 Accuracy

The results of three separate ANOVAs with Lag (3,9) and Group (Novice, Cellist) showed that when T2 was a human voice, the main effect of Group approached the margin of significance [$F_{(1, 34)} = 3.45, p = 0.07, \eta_p^2 = 0.092$], while both the main effect of Lag [$F_{(1, 34)} = 0.01, p = 0.91, \eta_p^2 = 0.00$] and Lag × Group interaction effect [$F_{(1, 34)} = 0.38, p = 0.54, \eta_p^2 = 0.011$] failed to reach statistical significance. A Bayesian ANOVA revealed that the data supported the null model 4.22 times over the model with Lag-only (BFInclusion = 0.237). This is moderate evidence in favor of the absence of the Lag effect. This means that when T2 is a human voice, T2|T1 accuracy rates are more likely to be lag-independent (i.e., no attentional blink effect for human voices). There was inconclusive evidence against the model including the Group as a factor, which suggests a weak change in the odds favoring the null model (BFInclusion = 0.846). Finally, there was an inconclusive evidence against the interaction model of Lag × Group (BFInclusion = 0.404).

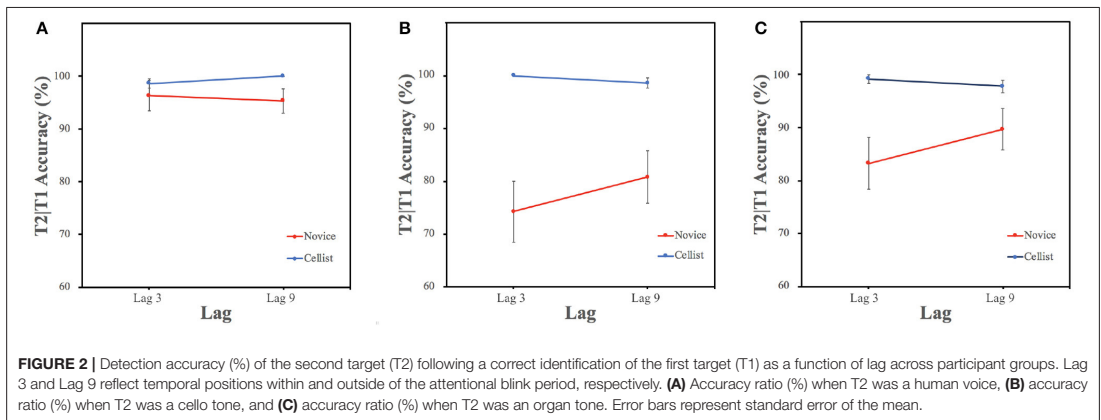
When T2 was a cello tone, the two-way ANOVA revealed a significant main effect of Group [$F_{(1, 34)} = 20.95, p < 0.001, \eta_p^2 = 0.381$], but the main effect of Lag [$F_{(1, 34)} = 0.968, p = 0.332, \eta_p^2 = 0.028$] and the Lag × Group interaction [$F_{(1, 34)} = 2.30, p = 0.189, \eta_p^2 = 0.063$] were non-significant. Based on the Bayesian ANOVA model comparisons, the data suggested extreme evidence favoring the Group-only model over the null model (BFInclusion = 365.803), suggesting the variances in the T2|T1 performance accuracy data was 365.803 times more likely to be explained with the main effect of Group over the absence of this effect. The model including Lag as the only factor indicated anecdotal evidence favoring the null model over the Lag-only model (BFInclusion = 0.395). The Lag × Group interaction model decreased the degree of support for the null model compared to the Lag-only model (BFInclusion = 0.687). This means that there was inconclusive evidence favoring the absence of the AB effect in the cello condition and that Group is the most likely candidate in explaining the differences in the T2|T1 performance compared to Lag and the interaction of the two.

When T2 was an organ tone, a two-way ANOVA indicated a significant main effect of Group [$F_{(1, 34)} = 8.279, p = 0.007, \eta_p^2 = 0.196$] and a significant Lag × Group interaction [$F_{(1, 34)} = 4.550, p = 0.040, \eta_p^2 = 0.118$]. The effect of Lag did not reach statistical significance, $F_{(1, 34)} = 1.891, p = 0.178, \eta_p^2 = 0.053$. The estimated Bayes factors indicated that the model received most support against the null model was the Group-only model (BFInclusion = 6.721). The variances in the T2|T1 accuracy data was 6.721 times more likely to be explained with the main effect of Group over the null model. The data suggested anecdotal evidence against the Lag-only model (BFInclusion = 0.529), meaning that there was a slight tendency toward the lag-independence of the T2|T1 performance in the organ condition but the evidence was inconclusive to support this claim further. Finally, there was only inconclusive evidence favoring the interaction model (Lag × Group), BFInclusion = 1.655. The results of two separate Bayes Paired samples *t*-tests with the T2|T1 performances at Lag 3 and Lag 9 showed that while there was a conclusive evidence for no difference for the organ tone detection within and outside the

TABLE 2 | Summary of the three-way ANOVA results on T2|T1 accuracy for all lag conditions using Traditional (reported with p -values) and Bayesian (reported with Bayes Factor estimations) statistics.

	Effect	Statistics	p -value	Bayes factor	Evidence category
Lag 1 vs. Lag 9	T2 type	$F_{(2, 68)} = 8.01$	$p < 0.001$	BF = 2035	Extreme evidence for H1
	Group	$F_{(1, 34)} = 13.37$	$p < 0.001$	BF = 34.05	Very strong evidence for H1
	T2 type \times group	$F_{(2, 68)} = 6.97$	$p = 0.002$	BF = 1,940.62	Extreme evidence for H1
	Lag	$F_{(1, 34)} = 0.29$	$p = 0.59$	BF = 0.153	Moderate evidence for H0
	Lag \times group	$F_{(1, 34)} = 0.35$	$p = 0.56$	BF = 0.210	Moderate evidence for H0
	T2 type \times lag	$F_{(1.75, 59.44)} = 1.58^a$	$p = 0.22$	BF = 0.186	Moderate evidence for H0
	Lag \times T2 type \times group	$F_{(1.75, 59.44)} = 1.60^a$	$p = 0.21$	BF = 0.305	Moderate evidence for H0
Lag 2 vs. Lag 9	T2 type	$F_{(1.76, 59.70)} = 9.89^a$	$p < 0.001$	BF = 399.39	Extreme evidence for H1
	Group	$F_{(1, 34)} = 17.59$	$p < 0.001$	BF = 84.10	Very strong evidence for H1
	T2 type \times group	$F_{(1.68, 57.07)} = 10.35^b$	$p < 0.001$	BF = 1,402.02	Extreme evidence for H1
	Lag	$F_{(1, 34)} = 0.39$	$p = 0.53$	BF = 0.160	Moderate evidence for H0
	Lag \times group	$F_{(1, 34)} = 0.39$	$p = 0.53$	BF = 0.244	Moderate evidence for H0
	T2 type \times lag	$F_{(1.36, 46.37)} = 0.15^b$	$p = 0.78$	BF = 0.099	Strong evidence for H0
	Lag \times T2 type \times group	$F_{(1.36, 46.37)} = 0.76^b$	$p = 0.43$	BF = 0.225	Moderate evidence for H0
Lag 3 vs. Lag 9	T2 type	$F_{(2, 68)} = 9.68$	$p < 0.001$	BF = 1,236.38	Extreme evidence for H1
	Group	$F_{(1, 34)} = 17.63$	$p < 0.001$	BF = 128.94	Extreme evidence for H1
	T2 type \times group	$F_{(2, 68)} = 9.66$	$p < 0.001$	BF = 4,807.10	Extreme evidence for H1
	Lag	$F_{(1, 34)} = 1.56$	$p = 0.22$	BF = 0.284	Moderate evidence for H0
	Lag \times group	$F_{(1, 34)} = 2.49$	$p = 0.12$	BF = 0.684	Anecdotal evidence for H0
	T2 type \times lag	$F_{(1.69, 57.62)} = 0.46^a$	$p = 0.60$	BF = 0.110	Moderate evidence for H0
	Lag \times T2 type \times group	$F_{(1.69, 57.62)} = 2.22^a$	$p = 0.12$	BF = 0.370	Anecdotal evidence for H0

^aHuynh-Feldt corrected, ^bGreenhouse-Geisser corrected.



blink periods in the cellist group (BF-0 = 0.117), the evidence remained inconclusive in the novice group (BF-0 = 1.823).

3.2.2. Group Differences on T2|T1 Accuracy

To further explore the group differences, two separate repeated measures ANOVAs with Lag (3,9) and T2 Type (voice, cello, organ) as well as the Bayesian equivalent of the same tests were conducted (see Table 3 for Bayesian results). In the novice

group, there was a significant main effect of T2 Type [$F_{(2, 34)} = 10.178$, $p < 0.001$, $\eta_p^2 = 0.374$] while the main effect of Lag and the interaction effect of T2 Type \times Lag were non-significant [$F_{(1, 17)} = 2.093$, $p = 0.166$, $\eta_p^2 = 0.110$ and $F_{(2, 34)} = 1.254$, $p = 0.298$, $\eta_p^2 = 0.069$, respectively]. For the same group, a Bayesian ANOVA revealed extreme evidence supporting the model with the T2 Type only (BF_{Inclusion} = 4,538.056), which means that the variances in the T2|T1 accuracy among

TABLE 3 | Analysis of effects in novices and in cellists obtained from two separate Bayesian ANOVAs with T2 Type and Lag as variables on T2|T1 performances.

Effects	P (incl)	P (incl data)	BF Incl
Analysis of effects in novices			
T2 type	0.400	0.921	4538.056
Lag	0.400	0.295	0.472
T2 type × lag	0.200	0.079	0.266
Analysis of effects in cellists			
T2 type	0.400	0.979	1213.098
Lag	0.400	0.229	0.306
T2 type × lag	0.200	0.021	0.090

Both analyses compare models that contain the effect to equivalent models stripped of the effect. Higher-order interactions are excluded.

the novices is roughly 4,538 times more likely to be caused by the T2 Type condition than the absence of this effect. The data also suggested anecdotal evidence favoring the null model over the Lag-only model and moderate evidence favoring the null model over the T2 Type × Lag interaction ($BF_{Inclusion} = 0.472$; $BF_{Inclusion} = 0.266$, respectively), meaning that the evidence is inconclusive for lag-dependence while there is moderate evidence against the combined effect of T2 Type and Lag on the T2|T1 accuracy data in this participant group.

In the cellist group, there was a borderline significant T2 Type × Lag interaction effect, $F_{(2, 34)} = 2.779$, $p = 0.076$, $\eta_p^2 = 0.141$. The main effect of T2 Type [$F_{(2, 34)} = 0.465$, $p = 0.632$, $\eta_p^2 = 0.027$] and the main effect of Lag [$F_{(1, 17)} = 0.619$, $p = 0.442$, $\eta_p^2 = 0.035$] were not statistically significant. For the cellists, a Bayesian ANOVA indicated extreme evidence in favor of the T2 Type over the null model ($BF_{Inclusion} = 1,213.098$). This means that the variances in the T2|T1 accuracy among the cellists is ~1,213 times more probable to occur due to T2 type condition than the absence of this effect. There was moderate evidence supporting the null model over the model with Lag as the only factor ($BF_{Inclusion} = 0.306$) and strong evidence favoring the null model over the T2 Type × Lag interaction model ($BF_{Inclusion} = 0.090$). This suggests a moderate support for the lag-independence and a strong support against the combined effect of Lag and T2 Type on the T2|T1 accuracy in this group. To test whether the cellists were better at detecting cello tones than the organ tones a Wilcoxon signed-rank test was performed. The results indicated a borderline statistically significant difference in detecting cello tones and organ tones for the expert cellists ($Z = 33.00$, $p = 0.042$, $r = -0.614$). Median correct detection rates of the cellists were 1.0 for both their own instrument and the organ. Bayes Factor analysis indicated only an anecdotal evidence for this difference, $BF_{10} = 1.403$.

Percentages of T2 hit and false alarm rates for T1-present trials (see **Table 4**) provide further insight into each groups' performance under different T2 conditions. For cellists, the false alarm rates were generally low (between 1 and 3%) across all T2 conditions. For novices, however, the false alarm rates varied across the different conditions. The rates were especially high under the cello (26%) and organ (16%) conditions, while under

TABLE 4 | Percentages of T2|T1 false alarms (FA) and hits.

	% Hit			% FA		
	Voice	Cello	Organ	Voice	Cello	Organ
Cellists	99.28	99.64	97.97	1.09	2.85	1.27
Novices	97.71	78.88	90.09	3.10	25.81	16.33

the voice condition the rates approached those of the cellists' (3% for the novices, 1% for the cellists). Thus, the results of the novices are highly likely to have been affected by response bias under the cello and organ conditions.

3.3. Correlation Analyses of the Maximal Attentional Blink Size and the Gold-MSI and LNS Scores

The maximal AB for each individual was calculated (in a similar fashion suggested by Colzato et al., 2007) by subtracting the minimum T2|T1 performance for each individual at an inside the blink period (whichever short lag yields the lowest performance) from the T2|T1 accuracy at the outside the blink period (Lag 9). Maximal AB scores from three cases were identified as outliers (Z -scores > 3.29) and were then assigned raw-scores one unit smaller so that they remain deviant but to a lesser extent, following the procedure recommended by Tabachnick and Fidell (2013). Bayesian correlation analyses were then conducted between the maximal AB data across T2 conditions and the Gold-MSI subscales as well as the LNS scores. All reported correlations are measured with Kendall's tau correlation coefficient. Results of these correlation analyses separately for each participant group are summarized in **Table 5** together with the mean values and standard deviations.

3.3.1. The Maximal AB and Gold-MSI Correlations

3.3.1.1. The maximal AB and musical sophistication factor

As the maximal AB reflects the largest temporal cost of selective attention for each individual (i.e., the bigger the value for the maximal AB, the larger the temporal cost), assuming that more musically sophisticated individuals would have suffered from the temporal costs to a lesser extent, negative correlations between the maximal AB and the musical sophistication score were expected. For all participants, negative correlations were supported between these two scores only in the cello condition, indicating moderate evidence (Kendall's tau = -0.266 , $BF_{-0} = 5.296$). The negative correlation between the musical sophistication scores and the maximal AB in the organ condition only had an anecdotal support (Kendall's tau = -0.227 , $BF_{-0} = 2.65$). The correlation analysis between the musical sophistication and the maximal AB in the voice condition, however, reflected a Bayes factor equaled 0.264 for a one-sided test where H_1 is specified as a negative correlation. This reflects a strong evidence for the null hypothesis. Thus, musical sophisticated levels ceased to be relevant in predicting the magnitude of the T2 deficit when T2 was a voice. As illustrated in **Figure 3**, individuals with relatively higher musical sophistication scores generally showed smaller

TABLE 5 | Mean scores of the Goldsmith Musical Sophistication inventory (Gold-MSI) and the Letter-Number Sequencing Task (LNS) and Kendall's tau correlation coefficient between the individual's maximal attentional blink sizes and the Gold-MSI and LNS scores.

	Novices			Cellists				
	Mean (SD)	Maximal AB voice	Maximal AB cello	Maximal AB organ	Mean (SD)	Maximal AB voice	Maximal AB cello	Maximal AB organ
Musical sophistication	55.67 (12.23)	0.13	0.06	0.18	108.5 (6.39)	0.40	0.06	-0.39*
Active engagement	27.83 (9.13)	0.21	-0.27	-0.06	40.67 (7.02)	0.33	0.30	-0.26
Perceptual abilities	39.11 (7.75)	0.24	0.08	0.04	58.56 (3.67)	0.09	-0.13	-0.48**
Musical training	10.78 (3.57)	0.10	0.12	0.16	43.94 (2.04)	0.42	0.22	-0.08
Singing abilities	23.89 (8.80)	-0.04	0.13	0.23	41.17 (3.28)	0.34	-0.04	-0.10
Emotions	31.28 (7.30)	0.15	-0.02	-0.04	36.00 (3.32)	-0.33*	-0.06	-0.18
LNS score	9.28 (2.78)	-0.15	-0.16	-0.20	10.22 (2.02)	-0.06	-0.37*	-0.24

For all tests alternative hypothesis specifies that the correlation is negative. Maximal AB = T2|T1Lag9 - T2|T1min. *BF > 3, **BF > 10. Bold values represent conclusive evidence for a negative correlation.

AB effect in the cello condition, while the maximal AB in the voice condition seems to have no connection with the individual's musical sophistication scores.

3.3.1.2. The maximal AB and the other Gold-MSI factors

The correlations between the maximal AB and the five other factors of the Gold-MSI were also explored using the Bayesian correlation matrix. There was strong evidence indicating a negative correlation between the maximal AB in the cello condition and active engagement factor (Kendall's tau = -0.311; BF-0 = 13.483), as well as moderate evidence for the negative correlations of the maximal AB for cello condition with the factors of perceptual abilities (Kendall's tau = -0.286; BF-0 = 7.878) and musical training (Kendall's tau = -0.274; BF-0 = 6.237). Similarly in the organ condition, moderate evidence supporting the negative correlations between maximal AB and active engagement (Kendall's tau = -0.252; BF-0 = 4.047) and perceptual abilities factors (Kendall's tau = -0.287; BF-0 = 8.145) was found. The strength of all the reported correlations were, however, weak. No other correlations reached a moderate support (BF-0 > 3) under a one-tailed test.

3.3.2. The Maximal AB and LNS Correlations

Bayesian correlation matrix indicated that the correlations between the LNS scores and maximal AB for the T2 Type conditions under the one-tailed test were inconclusive (BF-0 < 3). Although when examining the two participant groups separately, we observed that only the data obtained from cellists in the cello condition showed negative correlations between maximal AB and LNS (Kendall's tau = -0.372; BF-0 = 5.192). However, the strength of this negative correlation was weak.

4. DISCUSSION

The main purpose of the present study was to determine whether the temporal cost of selective allocation of auditory attention could be lowered or even be eliminated via manipulating the target objects relating to an expert's domain of expertise. Two types of expertise were considered to be interesting to this aim:

perceptual expertise with human voices (supposedly common to almost all of us) and musical expertise. The selection of the expert cellists in particular allowed us to test out not only a potential benefit relating to the trained instrument vs. a benefit for all auditory targets (near transfer of the training), but also to explore whether our common perceptual expertise with human voices could possibly have an extended attentional benefit for tones that are acoustically similar to voices (e.g., cello).

4.1. Voices Are Least Likely to Suffer From an Auditory AB Effect

The current study demonstrated that neither the instrumental tones of cello or organ nor human voices indicated evidence for the presence of an attentional blink effect, independent of one's musical expertise. The Bayesian analysis allowed us to compare the likelihood of this null effect. Although the evidence was inconclusive to argue for the lag-independence hypothesis (i.e., no support for a temporal cost in the T2 performance) for the cello and organ tones, the evidence supporting the lag-independence was larger (i.e., moderate evidence) in the human voice condition as compared with the cello and organ conditions, meaning that the human voices (when presented as the second target in the RAP) were the least likely to suffer from a temporal limitation to selective auditory attention. This is in line with the hypothesis that human voices as T2 would be less susceptible to suffer an attentional blink effect than the instrumental tones. There are a number of possible explanations that could be argued as to why human voices were the least likely to suffer from an AB effect. A straightforward explanation is that human voices have higher salience than the other target types and they therefore involuntarily capture attention. Hence they are more easily detected in comparison to other targets (i.e., instrumental tones) in this experiment. Weiss et al. (2016) provides physiological support for the heightened salience of voices compared to familiar instrumental timbres, with an observation of greater pupil dilation (reflecting enhanced arousal and stimulus salience) for vocal than for piano melodies. However, given that the AB task by its nature requires voluntary deployment of attention to the targets, the observed attentional

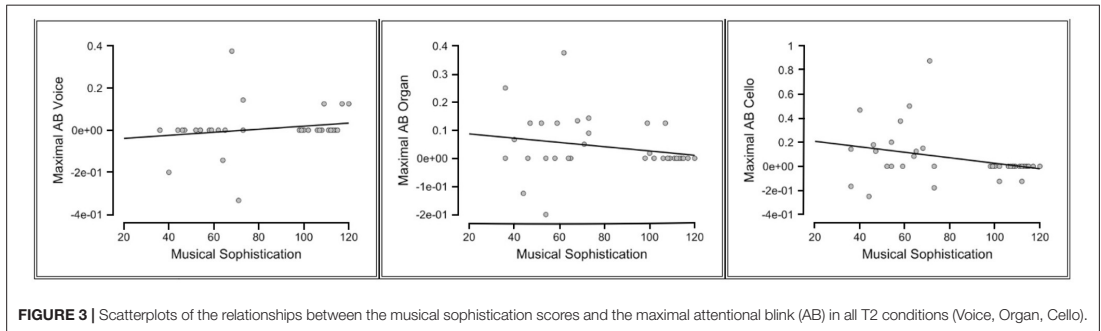


FIGURE 3 | Scatterplots of the relationships between the musical sophistication scores and the maximal attentional blink (AB) in all T2 conditions (Voice, Organ, Cello).

benefit for the selection of human voices cannot be explained with the bottom-up salience alone, but rather with an interaction of the top-down mechanisms and bottom-up influences on the temporal selective attention. The interaction between the top-down and bottom up biases is argued to increase the probabilistic competition in the favor of salient target for the object selection (Shinn-Cunningham, 2008), which in this case could explain a more successful detection of the human voices than the other target types. This explanation is consistent with the findings showing other salient targets, such as one's own name (Shapiro et al., 1997) are more likely to survive the attentional blink effect. From an evolutionary point of view, it could even be argued that voices are more salient than one's own name. There is evidence for a very early development of the processing mechanisms for voices and that even fetuses are capable of discriminate human voices from other sounds (Yovel and Belin, 2013). Indeed, it has been reported that the ability to respond to segmented speech develops as early as 36–40 weeks gestational age, and that fetuses can recognize maternal voice (Kisilevsky et al., 2003), which arguably is the most salient auditory stimuli of all due to its evolutionary importance.

The ease of detecting human voices, despite the temporal limitations of the human attentional system, could also be facilitated by the expertise that all humans share when it comes to the processing of human voices. This is in line with other expertise-related benefits that has been observed when the targets are drawn from any objects of expertise. In the sense that voices carry important information about the identity and emotional state of the speaker, they have been referred to in the literature as “auditory faces” (Belin et al., 2004). Using functional magnetic resonance imaging (fMRI), Belin et al. (2000) showed evidence for regions in the human brain that are strongly selective to human voices. These voice-selective areas in the superior temporal sulcus (STS) have been argued to potentially represent the auditory counterpart of the face-selective areas (FFA; Kanwisher et al., 1997) in the visual cortex. In concert with the well-established finding that there is no attentional blink for human faces, the long exposure and experience with the human voices is likely to make all humans capable of processing human voices with no temporal attentional cost. This explanation has support from the previous studies regarding the role of attention in human voice processing. For instance, Levy et al. (2003) found

that human voices elicit an ERP component related to Novelty P3 and P3a, suggesting an attentional capture.

It could also be argued that voices may be processed differently than other auditory objects. Beyond the AB paradigm, the overall accuracy rates (across all lag conditions) in the present study was the highest in the human voice condition, which may reflect a processing advantage for voices. In the literature, there are behavioral indications of a processing advantage for voices demonstrated by faster reaction times and lower duration thresholds (e.g., Agus et al., 2010; Agus et al., 2012) in detection and categorization of voices in comparison to the musical instrument tones, with an exception being that this advantage was only observed when RMS-level normalization was used as shown by Bigand et al. (2011). The data in the present study, despite the use of the peak normalization method, indicated a voice superiority in comparison to cello and organ tones reflected in the T2/T1 performance.

The lack of AB for voices may appear inconsistent with some of the findings in the literature using stimuli that also belong to the human voice category, such as spoken letters, spoken digits, and syllables. In the present study, the task required making a type distinction (i.e., reporting hearing a voice), while in the most other studies a token distinction (i.e., reporting what the voice said) was required. Furthermore, in difference from the present study, the distracters in those studies were also typically belonging to the human voice category (e.g., targets were spoken letters, distracters were spoken digits in Martens et al., 2015, and both targets and distracters were spoken syllables in Duncan et al., 1997 and in Tremblay et al., 2005), while in the present study distracters were environmental sounds. By inducing a pop-out effect for human voices presented among environmental sounds, the present study might have a lack of interference between the targets.

4.2. The Benefits of Musical Expertise in the Auditory AB Task

In the present study, expertise in music has shown clear benefits in both T1 and T2/T1 performance within the AB task. The current findings supported the lag-independence hypothesis for the expert cellists, as the data showed conclusive evidence for no difference between T2/T1 performances at the inside and outside of the blink periods. Importantly, T2 Type did not modulate

the absence of the AB within this group, which speaks in favor of a more generalized auditory attention benefit for the expert musicians in the current auditory AB paradigm.

In all conditions, as expected, the expert cellists had overall higher accuracy rates than the novices. These benefits (reflected by the evidence for group level differences in T2/T1 performance) were extreme when T2 was the cello and moderate when T2 was an organ tone but importantly, this advantage of the expert musicians disappeared when T2 was a human voice. There was a similar trend in the maximal AB scores as well, where the largest difference in the maximal AB scores between the cellists and novices was observed in the cello condition. These data suggested that expert cellists were more likely to have the T2/T1 benefits for their own instrument's tone and some benefit for other instrumental tones in comparison to the novices. This might have happened due to several reasons:

Being more experienced in regards to their principal instrument may have led to an enhancement of the processing of the tones of that instrument. However, the direct comparison of the differences in cello and organ tone detection rates in cellists only had anecdotal support. To put in another way, although the cellist showed the largest attentional benefits for their own instrument in comparison to the novices, the cellists were almost equally good at detecting organ tones and cello tones. Better allocation of auditory attention for cellists compared to the novices under the cello condition reflects that the cellist did not suffer from the processing costs for the cello as T2 as much as the novices did. Importantly, the highest false alarm rates were observed under the cello condition and the performance of the novices under the cello condition was most likely as a result of guessing. Furthermore, the cellists' auditory working memory span (measured by LNS) was negatively correlated with the processing costs for cello as T2, thus, it is possible that working memory could have a special role in temporal selective attending of the experts for their trained instrument. Motivational and emotional salience of the trained instrument and unintended priming effect (as the cellists knew they volunteered to a study where there was a need for cellist volunteers) may also have an impact on the cellists' performance under this condition.

Although the expert cellists' performance in this study was at the peak level for the detection of human voices just as the instrumental tones, the reason behind this particular finding was most likely beyond their musical expertise, but a rather generalized advantage for all humans, a perceptual expertise for the processing of human voices, since an equally good performance was observed in the novices. This was further supported by the correlation results between the T2 processing costs and the musical sophistication scores in the voice condition.

4.3. Anecdotal Evidence Against the AB Effect for Organ Tones

The finding of no AB (although only anecdotally supported) for organ tones in this study was somewhat unexpected. The lag-independence of organ tones was only conclusive in the cellists group, which may reflect a general musicianship advantage. What remains unclear is the reason why the absence of AB for organ

tones was inconclusive for the novice participants. Although the T2/T1 performances of the novices were lower at Lag 3 (inside the AB time-window) than at Lag 9 (outside the AB time-window) in organ condition, the slope of this function was not steep enough to indicate an AB effect (in similarity to the cello condition). This could mean that either T2 performance decrement at Lag 3 was too small or that the recovery of this impairment at Lag 9 was not large enough to be captured efficiently by this design. It is possible that the results could be influenced by the parameters used in this study. By altering these parameters, such as presentation rate, stimulus duration, T1 and distracter type, and task difficulty, perhaps it would be possible to obtain conclusive evidence for the temporal costs of attention for instrumental timbres in novices. Alternatively, increasing the power of the study by testing more participants could also help achieving conclusive evidence.

4.4. Limitations and Future Directions

The stimulus presentation rate in the present study (6.25 Hz) could be an important factor in explaining our findings. This rate could be too slow to produce a reliable auditory AB effect, as the presentation rate can have dramatic effects on the auditory AB magnitude (Arnell and Jolicoeur, 1999; Shen and Alain, 2010). This may have resulted in giving participants enough time to consolidate the targets into short-term memory. Alternatively, theta entrainment may be an underlying factor for not finding supporting evidence for the presence of an auditory AB effect in none of the conditions tested here. Recently, Shapiro et al. (2017) tested the effects of brain oscillations generated by the rate of the visual presentation stream on the T2 deficit, and found that the visual AB magnitude was the smallest at the theta range (6.26 Hz) in comparison to the alpha (10.3 Hz), beta (16.0 Hz), and gamma (36 Hz) frequency ranges. One future direction for research would be to explore whether an auditory AB effect would occur for voices and instrumental tones as T2 when the stimuli are presented at the rates of the alpha (10.3 Hz) and beta (16.0 Hz) oscillatory frequencies.

Another limitation of our study is that the T2 detection task might have been too easy to capture temporal cost of selective attention, especially in the expert performance. Here the task difficulty was not manipulated to trace individual thresholds, which could also explain the observed ceiling effect in the expert group. It is also possible that the nature of the T1 and T2 task in the design (with a 50–50 chance at obtaining a correct response) made the accuracy data more prone to ceiling effects. Future studies could manipulate the task difficulty based on individual performance and decrease the chance level for correct response to eliminate the potential confound of ceiling effects. Similarly, T1 discrimination task (noise vs. tone among environmental sounds) could also be too easy to produce a reliable auditory AB effect. Several studies investigated the link between T1 task difficulty and the AB magnitude through for example manipulating the effectiveness of the T1 backward masking (e.g., Seiffert and Lollo, 1997; Visser, 2007), response demands of the T1 task (e.g., Jolicoeur, 1999), or T1 perceptual load (e.g., Giesbrecht et al., 2009). Ouimet and Jolicoeur (2007) also argued that the data-limited difficulty manipulations (based on low-level perceptual qualities) may be less likely to have

a modulatory effect on the AB than the resource-limited T1 manipulations, depending on the impact of the changes in the duration of central processing for this task. So manipulating different factors affecting T1 task difficulty might alter the AB results we have observed in this study.

Finally, as the expert musician population consisted exclusively of cellists, it is difficult to generalize the conclusions to other expert musician groups. However, thanks to this restriction in the participant group, it was possible to contrast the tones of an expert musician's instrument of expertise vs. other instrumental tones that they have not been trained on.

5. CONCLUSION

Under the parameters used in this study, human voices seem to be less susceptible to the auditory AB effects, independent of musical expertise. However, this general effect does not seem to extend to cello tones despite its perceptual similarity to human voices. If this were to be true, we would have observed no group differences. On the contrary, the largest group difference in T2/T1 performance was observed in the cello condition, which reflected an attentional advantage for the expert cellists when the second target was the tones of their principal instrument more than for the other target conditions. Lastly, experts had an overall benefit in the AB task, reflected in both T1 and T2/T1 performances, the exception being the human voice condition.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/[Supplementary Material](#).

REFERENCES

- Agus, T. R., Sued, C., Thorpe, S. J., and Pressnitzer, D. (2012). Fast recognition of musical sounds based on timbre. *J. Acoust. Soc. Am.* 131, 4124–4133. doi: 10.1121/1.3701865
- Agus, T. R., Thorpe, S. J., Sued, C., and Pressnitzer, D. (2010). "Characteristics of human voice processing," in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems* (Paris: IEEE), 509–512.
- Akyürek, E. G., BernhardHommel, and Jolicoeur, P. (2007). Direct evidence for a role of working memory in the attentional blink. *Mem. Cogn.* 35, 621–627. doi: 10.3758/BF03193300
- Arnell, K. M., and Jenkins, R. (2004). Revisiting within-modality and cross-modality attentional blinks: Effects of target–distractor similarity. *Percept. Psychophys.* 66, 1147–1161. doi: 10.3758/BF03196842
- Arnell, K. M., and Jolicoeur, P. (1999). The attentional blink across stimulus modalities: evidence for central processing limitations. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 630–648. doi: 10.1037//0096-1523.25.3.630
- Arnell, K. M., and Larson, J. M. (2002). Cross-modality attentional blinks without preparatory task-set switching. *Psychon. Bull. Rev.* 9, 497–506. doi: 10.3758/BF03196305
- Askenfelt, A. (1991). "Voices and strings: close cousins or not?" in *Music, Language, Speech and Brain*, eds J. Sundberg, L. Nord, and R. Carlson (London: Palgrave), 243–256.
- Awh, E., Serences, J., Laurey, P., Dhaliwal, H., Jagt, T. V., and Dassonville, P. (2004). Evidence against a central bottleneck during the attentional blink:

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Norwegian Centre for Research Data (NSD) on 13 February 2018 [ref. no: 58784/3]. The study was performed in accordance with the recommendations of the Declaration of Helsinki. The patients/participants provided their written informed consent to participate in this study

AUTHOR CONTRIBUTIONS

MA conceived and designed the study. MA and BL performed the statistical analyses. MA wrote the first draft of the manuscript. RG and BL edited the parts of the manuscript. All authors contributed to the manuscript revision, read, and approved the submitted version.

FUNDING

This work was partially supported by the Research Council of Norway through its Centers of Excellence scheme, project number 262762.

ACKNOWLEDGMENTS

We thank RITMO Center of Excellence for providing us with an external research fund for the open access publication fee.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02935/full#supplementary-material>

Multiple channels for configural and featural processing. *Cogn. Psychol.* 48, 95–126. doi: 10.1016/S0010-0285(03)00116-6

Belin, P., Fecteau, S., and Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends Cogn. Sci.* 8, 129–135. doi: 10.1016/j.tics.2004.01.008

Belin, P., Zatorre, R., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature* 40, 309–312. doi: 10.1038/35002078

Bigand, E., Delbé, C., and Tillmann, Y. G. B. (2011). Categorization of extremely brief auditory stimuli: domain-specific or domain-general processes? *PLoS ONE* 6:e27024. doi: 10.1371/journal.pone.0027024

Blacker, K. J., and Curby, K. M. (2016). The attentional blink is attenuated for objects of expertise. *Vis. Cogn.* 24, 132–140. doi: 10.1080/13506285.2016.1200701

Chartrand, J., and Belin, P. (2006). Superior voice timbre processing in musicians. *Neurosci. Lett.* 405, 164–167. doi: 10.1016/j.neulet.2006.06.053

Chartrand, J., Peretz, I., and Belin, P. (2008). Auditory recognition expertise and domain specificity. *Brain Res.* 1220, 191–198. doi: 10.1016/j.brainres.2008.01.014

Colzato, L. S., Spape, M. M. A., Pannebakker, M. M., and Hommel, B. (2007). Working memory and the attentional blink: blink size is predicted by individual differences in operation span. *Psychon. Bull. Rev.* 14, 1051–1057. doi: 10.3758/BF03193090

Diens, Z. (2014). Using bayes to get the most out of non-significant results. *Front. Psychol.* 5:781. doi: 10.3389/fpsyg.2014.00781

Duncan, J., Martens, S., and Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature* 387, 808–810. doi: 10.1038/42947

- Gaser, C., and Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *J. Neurosci.* 23, 9240–9245. doi: 10.1523/JNEUROSCI.23-27-09240.2003
- Giesbrecht, B., Sy, J. L., and Lewis, M. K. (2009). Personal names do not always survive the attentional blink: behavioral evidence for a flexible locus of selection. *Vis. Res.* 49, 1378–1388. doi: 10.1016/j.visres.2008.02.013
- Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Ewans, A. C., et al. (2009). Musical training shapes structural brain development. *J. Neurosci.* 29, 3019–3025. doi: 10.1523/JNEUROSCI.5118-08.2009
- JASPTeam (2018). *JASP (Version 0.8.6)* [Computer Software]. Amsterdam.
- Jeffreys, H. (1961). *The Theory of Probability*. Oxford: OUP.
- Johnston, S. J., Linden, D. E. J., and Shapiro, K. L. (2012). Functional imaging reveals working memory and attention interact to produce the attentional blink. *J. Cogn. Neurosci.* 24, 28–38. doi: 10.1162/jocn_a_00054
- Jolicoeur, P. (1999). Concurrent response-selection demands modulate the attentional blink. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 1097–1113. doi: 10.1037/0096-1523.25.4.1097
- Kanwisher, N., McDermott, N., and Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302–4311. doi: 10.1523/JNEUROSCI.17-11-04302.1997
- Kisilevsky, B. S., Hains, S. M., Lee, K., Xie, X., Huang, H., Ye, H. H., et al. (2003). Effects of experience on fetal voice recognition. *Psychol. Sci.* 14, 220–224. doi: 10.1111/1467-9280.02435
- Kraus, N., and Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat. Rev. Neurosci.* 11:599. doi: 10.1038/nrn2882
- Landau, A. N., and Bentin, S. (2008). Attentional and perceptual factors affecting the attentional blink for faces and objects. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 818–830. doi: 10.1037/0096-1523.34.4.818
- Lee, M. D., and Wagenmakers, E.-J. (2014). *Bayesian Cognitive Modeling: A Practical Course*. Cambridge: Cambridge University Press.
- Levy, D. A., Granot, R., and Bentin, S. (2003). Neural sensitivity to human voices: Erp evidence of task and attentional influences. *Psychophysiology* 40, 291–305. doi: 10.1111/1469-8986.00031
- Martens, S., Dun, M., Vries, M. D., and Smid, H. G. (2015). Musical minds: attentional blink reveals modality-specific restrictions. *PLoS ONE* 10:e118294. doi: 10.1371/journal.pone.0118294
- Martens, S., and Johnson, A. (2009). Working memory capacity, intelligence, and the magnitude of the attentional blink revisited. *Exp. Brain Res.* 192, 43–52. doi: 10.1007/s00221-008-1551-1
- Mathôt, S. (2017). *Bayes Like a Baw: Interpreting Bayesian Repeated Measures in JASP*.
- Moreno, S., and Bidelman, G. M. (2014). Examining neural plasticity and cognitive benefit through the unique lens of musical training. *Hear. Res.* 308, 84–97. doi: 10.1016/j.heares.2013.09.012
- Müllensiefen, D., Gingras, B., Musil, J., and Stewart, L. (2014). The musicality of non-musicians: an index for assessing musical sophistication in the general population. *PLoS ONE* 9:e89642. doi: 10.1371/journal.pone.0089642
- Opolko, F., and Wapnick, J. (2006). *The McGill University Master Samples Collection on DVD (3 DVDs)*. Quebec, QC: McGill University.
- Ouimet, C., and Jolicoeur, P. (2007). Beyond task 1 difficulty: the duration of t1 encoding modulates the attentional blink. *Vis. Cogn.* 15, 290–304. doi: 10.1080/13506280600693741
- Pezzuti, L., and Rossetti, S. (2017). Number sequencing, figure weights, and cancellation subsets of wais-iv administered to elders. *Pers. Individ. Differ.* 104, 352–356. doi: 10.1016/j.paid.2016.08.019
- Potter, M. C., Chun, M. M., Banks, B. S., and Muckenhoupt, M. (1998). Two attentional deficits in serial target search: the visual attentional blink and an amodal task-switch deficit. *J. Exp. Psychol. Learn. Mem. Cogn.* 24:979. doi: 10.1037/0278-7393.24.4.979
- Raymond, J. E., Shapiro, K. L., and Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: an attentional blink? *J. Exp. Psychol. Hum. Percept. Perform.* 18, 849–860. doi: 10.1037/0096-1523.18.3.849
- Segado, M., Hollinger, A., Thibodeau, J., Penhune, V., and Zatorre, R. (2018). Partially overlapping brain networks for singing and cello playing. *Front. Neurosci.* 12:351. doi: 10.3389/fnins.2018.00351
- Seiffert, A. E., and Lollo, V. D. (1997). Low-level masking in the attentional blink. *J. Exp. Psychol. Hum. Percept. Perform.* 23, 1061–1073. doi: 10.1037/0096-1523.23.4.1061
- Shapiro, K. L., Caldwell, J., and Sorensen, R. E. (1997). Personal names and the attentional blink: a visual cocktail party effect. *J. Exp. Psychol. Hum. Percept. Perform.* 23:504. doi: 10.1037/0096-1523.23.2.504
- Shapiro, K. L., Hanslmayr, S., Enns, J. T., and Lleras, A. (2017). Alpha, beta: the rhythm of the attentional blink. *Psychon. Bull. Rev.* 24, 1862–1869. doi: 10.3758/s13423-017-1257-0
- Shelton, J. T., Elliott, E. M., Hill, B. D., Calamia, M. R., and Gouvier, W. D. (2009). A comparison of laboratory and clinical working memory tests and their prediction of fluid intelligence. *Intelligence* 37, 283–293. doi: 10.1016/j.intell.2008.11.005
- Shen, D., and Alain, C. (2010). Neuroelectric correlates of auditory attentional blink. *Psychophysiology* 47, 184–191. doi: 10.1111/j.1469-8986.2009.00924.x
- Shen, D., and Alain, C. (2012). Implicit temporal expectation attenuates auditory attentional blink. *PLoS ONE* 7:e36031. doi: 10.1371/journal.pone.0036031
- Shen, D., and Mondor, T. A. (2006). Effect of distractor sounds on the auditory attentional blink. *Percept. Psychophys.* 68, 228–243. doi: 10.3758/BF03193672
- Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends Cogn. Sci.* 12, 182–186. doi: 10.1016/j.tics.2008.02.003
- Slawinski, E. B., Goddard, K., and Bonfield, S. (2002). Are musicians listening to sounds in a different way than non-musicians? *Can. Acoust.* 30, 162–163. Retrieved from: <https://jcaa.caa-aca.ca/index.php/jcaa/article/view/1505>
- Slevc, L. R., Davey, N. S., Buschkuhl, M., and Jaeggi, S. M. (2016). Tuning the mind: exploring the connections between musical ability and executive functions. *Cognition* 152, 199–211. doi: 10.1016/j.cognition.2016.03.017
- Tabachnick, B. G., and Fidell, L. S. (2013). *Using Multivariate Statistics, 6th Edn.* New Jersey, NJ: Pearson.
- Tremblay, S., Vachon, F., and Jones, D. M. (2005). Attentional and perceptual sources of the auditory attentional blink. *Percept. Psychophys.* 67, 195–208. doi: 10.3758/BF03206484
- Visser, T. A. (2007). Masking t1 difficulty: processing time and the attentional blink. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 285–297. doi: 10.1037/0096-1523.33.2.285
- Wechsler, D. (2008). *Manual for the Wechsler Adult Intelligence Scale, 4th Edn.* San Antonio, TX: Pearson.
- Weiss, M. W., Trehub, S. E., Schellenberg, E. G., and Habashi, P. (2016). Pupils dilate for vocal or familiar music. *J. Exp. Psychol. Hum. Percept. Perform.* 42, 1061–1065. doi: 10.1037/xhp0000226
- Werner, P., Swope, A., and Heide, F. J. (2006). The music experience questionnaire: development and correlates. *J. Psychol.* 140, 329–345. doi: 10.3200/JRPL.140.4.329-345
- Willems, C., and Martens, S. (2016). Time to see the bigger picture: individual differences in the attentional blink. *Psychon. Bull. Rev.* 25, 1289–1299. doi: 10.3758/s13423-015-0977-2
- Yovel, G., and Belin, P. (2013). A unified coding strategy for processing faces and voices. *Trends Cogn. Sci.* 17, 263–271. doi: 10.1016/j.tics.2013.04.004

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

Copyright © 2020 Akça, Laeng and Godoy. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Paper II

Recognition of brief sounds in rapid serial auditory presentation

**Merve Akça, Jonna Katariina Vuoskoski, Bruno Laeng, &
Laura Bishop**

Submitted for publication.



Paper III

What does the pupil whisper? Tracing the temporal costs of auditory selective attention with pupillometry

**Merve Akça, Laura Bishop, Jonna Katariina Vuoskoski, &
Bruno Laeng**

Under review in *Frontiers in Neuroscience*.



Appendices

Appendix A

Bayesian correlational tests for Papers II and III

Table A.1: Bayesian Correlation Table for Paper II Experiment 1a

			Pearson	
			BF ₁₀	BF ₀₁
Musical sophistication	-	d'	0.29	3.39
	-	RT	0.35	2.87
	-	Certainty level	0.30	3.34
d'	-	RT	0.44	2.29
	-	Certainty level	3.00	0.33
RT	-	Certainty level	4.49	0.22

Note. Bayes Factors (BF) are calculated using JASP's default priors. BF₁₀ indicates the likelihood of data given H1 compared to H0, while BF₀₁ indicates likelihood of data given H0 compared to H1.

Table A.2: Bayesian Correlation Table for Paper II Experiment 1b

			Pearson		Kendall	
			BF ₁₀	BF ₀₁	BF ₁₀	BF ₀₁
Musical sophistication	-	d'	7.35	0.17	-	-
	-	Digit span	0.34	2.93	-	-
	-	RT	0.26	3.84	-	-
	-	Certainty level	0.27	3.71	-	-
d'	-	Digit span	0.33	3.03	-	-
	-	RT	0.79	1.27	-	-
	-	Certainty level	0.25	3.92	-	-
Digit span	-	RT	-	-	0.38	2.61
	-	Certainty level	0.31	0.321	-	-
RT	-	Certainty level	0.41	2.45	-	-

Note. Bayes Factors (BF) are calculated using JASP's default priors. BF₁₀ indicates the likelihood of data given H1 compared to H0, while BF₀₁ indicates likelihood of data given H0 compared to H1. Kendall's tau correlation coefficient were used whenever the assumption of bivariate normality was violated.

A. Bayesian correlational tests for Papers II and III

Table A.3: Bayesian Correlation Table for Paper III

			Pearson		Kendall	
			BF ₁₀	BF ₀₁	BF ₁₀	BF ₀₁
Active engagement	-	ABmag	0.17	5.94	-	-
	-	Pupils 30 ms	0.19	5.15	-	-
	-	Pupils 90 ms	0.17	5.97	-	-
Perceptual abilities	-	ABmag	0.17	5.90	-	-
	-	Pupils 30 ms	0.17	5.77	-	-
	-	Pupils 90 ms	0.22	4.55	-	-
Musical training	-	ABmag	0.17	5.97	-	-
	-	Pupils 30 ms	0.17	5.95	-	-
	-	Pupils 90 ms	0.21	4.86	-	-
Singing abilities	-	ABmag	0.19	5.25	-	-
	-	Pupils 30 ms	0.22	4.51	-	-
	-	Pupils 90 ms	0.19	5.37	-	-
Emotions	-	ABmag	0.20	5.03	-	-
	-	Pupils 30 ms	0.22	4.44	-	-
	-	Pupils 90 ms	0.21	4.80	-	-
Musical sophistication	-	ABmag	0.17	5.95	-	-
	-	Pupils 30 ms	0.18	5.68	-	-
	-	Pupils 90 ms	0.21	4.79	-	-
Dysfunctional impulsivity	-	ABmag	-	-	0.37	2.72
	-	Pupils 30 ms	-	-	0.77	1.29
	-	Pupils 90 ms	-	-	0.94	1.06
Functional impulsivity	-	ABmag	0.23	4.33	-	-
	-	Pupils 30 ms	0.17	5.79	-	-
	-	Pupils 90 ms	0.17	5.85	-	-

Note. BF₁₀ indicates the likelihood of data given H1 compared to H0, while BF₀₁ indicates likelihood of data given H0 compared to H1. Dysfunctional impulsivity correlations are calculated with Kendall's tau coefficient as these scores violated the assumption of normality. ABmag refers to the magnitude of attentional blink.

Appendix B

Bayes Factor evidence categories

The following table shows how Bayes Factors were interpreted in the papers in this thesis.

Table B.1. Categorical classification scheme for the Bayes factor interpretation suggested by Lee and Wagenmakers (2013); adjusted from Jeffreys (1961).

Bayes factor	Evidence category
> 100	Extreme evidence for H1
$30 - 100$	Very strong evidence for H1
$10 - 30$	Strong evidence for H1
$3 - 10$	Moderate evidence for H1
$1 - 3$	Anecdotal evidence for H1
1	No evidence
$1/3 - 1$	Anecdotal evidence for H0
$1/10 - 1/3$	Moderate evidence for H0
$1/30 - 1/10$	Strong evidence for H0
$1/100 - 1/30$	Very strong evidence for H0
$< 1/100$	Extreme evidence for H0