

Prosody and attention orienting

The role of rising intonation in
speech processing

Maria Lialiou

Studies in Laboratory Phonology 16





Studies in Laboratory Phonology

Chief Editor: Martine Grice

Editors: Doris Mücke, Taehong Cho

In this series:

7. Tilsen, Sam. Syntax with oscillators and energy levels.
8. Ben Hedia, Sonia. Gemination and degemination in English affixation: Investigating the interplay between morphology, phonology and phonetics.
9. Easterday, Shelece. Highly complex syllable structure: A typological and diachronic study.
10. Roessig, Simon. Categoriality and continuity in prosodic prominence.
11. Schmitz, Dominic. Production, perception, and comprehension of subphonemic detail: Word-Final /s/ in English.
12. Schubö, Fabian, Sabine Zerbian, Sandra Hanne & Isabell Wartenburger (eds.). Prosodic boundary phenomena.
13. Albert, Aviad. A model of sonority based on pitch intelligibility.
14. Wehrle, Simon. Conversation and intonation in autism: A multi-dimensional analysis.
15. Kaland, Constantijn. Word stress in prosodic theory.
16. Lialiou, Maria. Prosody and attention orienting: The role of rising intonation in speech processing.

ISSN: 2363-5576

Prosody and attention orienting

The role of rising intonation in
speech processing

Maria Lialiou



Maria Lialiou. 2025. *Prosody and attention orienting: The role of rising intonation in speech processing* (Studies in Laboratory Phonology 16). Berlin: Language Science Press.

This title can be downloaded at:

<http://langsci-press.org/catalog/book/516>

© 2025, Maria Lialiou

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):

<http://creativecommons.org/licenses/by/4.0/> 

The work presented in this book is based on the author's doctoral dissertation, which was accepted by the Faculty of Arts and Humanities at the University of Cologne in 2024.

ISBN: 978-3-96110-511-3 (Digital)

978-3-98554-140-9 (Hardcover)

ISSN: 2363-5576

DOI: 10.5281/zenodo.15396923

Source code available from www.github.com/langsci/516

Errata: paperhive.org/documents/remote?type=langsci&id=516

Cover and concept of design: Ulrike Harbort

Typesetting: Hannah Schleupner

Proofreading: Cesar Perez Guarda, Daniela Schroeder, Elliott Pearl, Giorgia

Troiani, Jeroen van de Weijer, Kirill Maslinsky, Luciana Marques, Maria

Onoeva, Mary Ann Walter, Nicole Benker, Jean Nitzke

Fonts: Libertinus, Arimo, DejaVu Sans Mono

Typesetting software: Xe_{La}TeX

Language Science Press

Scharnweberstraße 10

10247 Berlin, Germany

<http://langsci-press.org>

support@langsci-press.org

Storage and cataloguing done by FU Berlin

Contents

Acknowledgments	v
Abbreviations	vii
1 General introduction	1
1.1 Topic, research questions, and main findings	1
1.2 Research rationale	6
1.3 Outline and preview of the next chapters	8
2 Attention in the auditory realm	11
2.1 What is attention orienting?	13
2.2 Involuntary and voluntary attention orienting	16
2.3 Auditory processing stages and their relation to attention	19
2.4 The role of rises in attention orienting	21
2.5 Approaches to studying attention orienting	25
2.5.1 Behavioural measures	25
2.5.2 Traditional psychophysiological methods	26
2.5.3 Brain measurements: The event-related potentials technique	26
2.5.4 ERP correlates of attention orienting: MMN and P3	30
2.5.5 Indexing attention orienting using pupillometry	33
2.6 Summary	36
3 Intonation and its association with prominence	39
3.1 Intonational phonology and prosodic structure	41
3.2 Intonational rises as a prominence-cueing device	45
3.3 Intonational rises in speech processing	48
3.4 Signal- and context-driven processing	53
3.5 A brief note on German list intonation	55
3.6 Summary	56
4 Unravelling the role of intonation in attention orienting using ERPs	59
4.1 Motivation	59

Contents

4.2	Methods	61
4.2.1	Participants	62
4.2.2	Speech materials	62
4.2.3	EEG recording	67
4.2.4	Procedure	68
4.2.5	Data pre-processing	70
4.2.6	Post-processing and statistical data analysis	71
	4.2.6.1 Post-processing	71
	4.2.6.2 Inference criteria	72
4.3	Results	73
4.3.1	Accentual contours	75
	4.3.1.1 Accentual falling deviant vs. accentual rising standard	75
	4.3.1.2 Accentual rising deviant vs. accentual falling standard	76
4.3.2	Boundary contours	77
	4.3.2.1 Boundary falling deviant vs. boundary rising standard	77
	4.3.2.2 Boundary rising deviant vs. boundary falling standard	79
4.3.3	Interim summary	79
4.4	Discussion	82
4.4.1	The processing of rises and falls and their contribution to attention orienting	82
4.4.2	The role of the phonological status of the rise (and the fall) in attention orienting	87
4.4.3	Some implications for the neural architecture of language	90
4.5	Summary	91
5	Indexing the attention orienting function of domain-final rises using pupillometry	93
5.1	Motivation	93
5.2	Methods	98
5.2.1	Participants	98
5.2.2	Speech materials	98
5.2.3	Acoustic characterisation of speech material	101
5.2.4	Experimental procedures	103
	5.2.4.1 Pupillometry	104

5.2.4.2	Cognitive test battery	105
5.2.4.2.1	Flanker task	106
5.2.4.2.2	Odd-man-out task	107
5.2.4.2.3	Digit span task	109
5.2.5	Data processing and statistical analyses	110
5.2.6	Data pre-processing	110
5.2.6.1	Pupil data	110
5.2.6.2	Cognitive test battery	111
5.2.6.3	Inference criteria	114
5.2.6.4	Confirmatory analysis	114
5.2.6.5	Exploratory analyses	115
5.2.6.5.1	Sequence length	115
5.2.6.5.2	Cognitive variability	116
5.3	Results	117
5.3.1	Confirmatory analysis	118
5.3.2	Exploratory analyses	119
5.3.2.1	Sequence length	119
5.3.2.2	Cognitive variability	121
5.3.3	Interim summary	124
5.4	Discussion	124
5.4.1	Rising intonation at the edges of constituents as an orienting device	125
5.4.2	The relevance of sequence length for attention orienting in language	129
5.4.3	The contribution of individual variability to attention orienting	131
5.5	Summary	135
6	General discussion	139
6.1	Summary and main observations	139
6.2	General implications of main findings	141
6.2.1	The relevance of intonation for the architecture of attention orienting in spoken language	141
6.2.2	The cognitive & functional contribution of intonational rises to attention orienting	145
6.2.3	The role of expectations in attention orienting	147
6.2.4	Neural and pupillary underpinnings of attention orienting	150
6.2.5	Individual variability and other top-down operations . .	151
6.3	Limitations and directions for future work	155

Contents

6.4	Conclusion	157
Appendix A: Pilot study – Button press		159
A.1	Participants	159
A.2	Speech materials	159
A.3	Experimental procedures	159
A.3.1	Button press	159
A.3.1.1	Cognitive test battery	160
A.4	Data processing and statistical analyses	160
A.4.1	Data pre-processing	161
A.4.1.1	Button press task	161
A.4.1.2	Cognitive test battery	161
A.4.2	Inference criteria	162
A.4.2.1	Confirmatory analyses	162
A.4.2.2	Exploratory analyses	162
A.4.2.2.1	Sequence length	162
A.4.2.2.2	Cognitive variability	163
A.4.2.2.3	Speed–accuracy trade-off	163
A.5	Results	163
A.5.1	Confirmatory analyses	164
A.5.2	Exploratory analyses	165
A.5.2.1	Sequence length	165
A.5.2.2	Cognitive variability	165
A.5.2.3	Speed–accuracy trade-off	169
A.6	Discussion	171
References		175
Index		197
Name index		197

Acknowledgments

This book is a revised version of my doctoral dissertation which was accepted by the Faculty of Arts and Humanities of the University of Cologne in 2024. The original acknowledgements are reprinted below with minor updates.

First and foremost, I would like to express my sincere gratitude to my two advisors, Petra Schumacher and Martine Grice, for being models of two fierce female researchers who guided me and inspired me all along. Petra and Martine are two unique individuals, admirable scientists, and excellent professors with a profound interest in their students. I would like to particularly thank both of them for embracing me in the team, persistently supporting me, questioning me, challenging me, protecting me, believing in me, and granting me freedom in pursuing my own scientific ideas! Petra and Martine have inspired me not only with their research ideas and their inquisitive characters, but also with their constant effort to push against boundaries and obstacles that early researchers are facing. I wish to become as fierce in pushing for support and equality in academia and throughout life. I would also like to thank Stefan Baumann, my third examiner, for always giving me detailed and valuable feedback.

Further, I want to acknowledge support from the DFG-funded CRC 1252 “Prominence in Language” which provided the institutional structure and funding for conducting my PhD. This work would not have been possible without the doctoral research position I have had the privilege to acquire in the project A01. Further, I thank the a.r.t.e.s. Graduate School for the Humanities Cologne for awarding me the Katharina Niemeyer Grant for language editing.

I would also like to thank all my wonderful colleagues from the “House of Prominence” for creating a very friendly environment, for sharing ideas and experiences, and for the fruitful – as well as bizarre – discussions we often had during lunch break. Big thanks goes to my office-buddy Janne Lorenzen, Malin Spaniol, Magdalena Repp, and Guendalina Reul who except for amazing colleagues and peers have become truly caring friends, supporting me in various aspects of my life. Moreover, I would like to thank Sandra Vella and Elina Savino for their support, mentoring, help, and advice. Additionally, I thank Christine Röhr and Aviad Albert for their invaluable input regarding my research, their support, and collaborative spirit. Further, I want to express my truthful gratitude

Acknowledgments

to Simon Wehrle for editing this entire dissertation. Big thanks goes to Ingmar Brilmayer as well, for teaching me how to run EEG experiments and how to process EEG data. I am also thankful to Jesse Harris for our collaboration, and for teaching me about pupillometry. Finally, I would like to thank Solveigh Janzen and Nadia Pelageina who supported me with data annotations and participant recruitment.

I would also like to thank all the other people (current and former members) from UoC who have influenced my work in many different, direct or indirect, ways. First of all, I want to thank my a.r.t.e.s. class 8 cohort. From the Department of German Language and Literature I, I want to thank Ina Bornkessel-Schlesewsky, Hanna Buhl, Paul Compensis, Claudia Kilter, Anne Lützeler, Umesh Patil, Clare Patterson, Brita Rietdorf, Clarissa Selegrad, Sophie Sprengel, Caterina Ventura, and Robert Voigt. Brita and Claudia deserve a special thank you for their enormous help with recruiting participants and running my experiments. From the Institute of Phonetics, I thank Florence Bails, Anna Bruggeman, Francesco Cangemi, T. Mark Elison, Alicia Janz, Constantijn Kaland, Theo Klinker, Lena Pagel, Simon Roessig, and Simona Sbranna.

Outside of Cologne, I would like to thank Shravan Vasishth, Bruno Nicenboim, Anna Laurinavichyute, and Bodo Winter for introducing me to the world of Bayesian thinking, and Márton Sóskuthy for teaching me the beauty of GAMMs. They have all shaped my profound interest in data analysis and open science. I am deeply grateful for the opportunity they gave me to consult with them about the analyses reported in this dissertation, acquiring invaluable feedback. I would also like to thank Yiya Chen for all the stimulating discussions we had about my EEG investigation, and Kalliopi Katsika for helping me navigate academia.

Besides academia, I would like to express my gratitude to my beloved parents, Theofilos and Tonia, and to my two amazing brothers, Stefanos and Pericles, for being there for me in every step of my life! Σας ευχαριστώ για όλα! Further, a big thanks to my beloved friends Maria and Evi for their endless support despite the geographical distance. Finally, to my wonderful partner and best friend: Bill, thanking you here is definitely not representative of the role you have played in my PhD endeavour and in my life in general. I cannot express enough how grateful I am for your never-ending patience and support. Thank you for agreeing to move in Cologne with me, thank you for encouraging me, believing in me, taking care of me when I had no time for anything else than working on this dissertation, listening to my ideas regardless of how bored you got, and cheering me up when I got overwhelmed. Thank you for loving me! Ως τον ουρανό και πάλι πίσω!

Abbreviations

AM	Autosegmental-metrical theory	MEG	Magnetoencephalography
ABR	Auditory brainstem response	min	Minute(s)
ACC	Anterior cingulate cortex	MLR	Middle-latency responses
CI	Confidence interval	MMN	Mismatch negativity
CNS	Central nervous system	ms	Millisecond(s)
CoG	Center of gravity	OR	Orienting response
CoM	Center of mass	OSF	Open science framework
CrI	Credible interval	PDR	Pupil dilation response
CSR	Central sound representation	PET	Positron emission tomography
CV	Consonant-vowel	SC	Superior colliculus
EEG	Electroencephalography	SPECT	Single photon emission computed tomography
ERP	Event-related potential	ToBI	Tones and breaks indices
f ₀	Fundamental frequency	TWI	Temporal window of integration
fMRI	Functional magnetic resonance imaging	WMC	Working memory capacity
GAMM(s)	Generalised additive mixed model(s)	WM	Working memory
GSR	Galvanic skin response	Δf_0	Delta f ₀
GToBI	German tones and breaks indices	μV	Microvolt
Hz	Hertz	σ	Syllable
IP	Intonational phrase	Σ	Foot
ip	Intermediate phrase	φ	Accentual phrase
LC	Locus coeruleus	ω	Prosodic word
L1	A speaker's first language		

1 General introduction

The world we live in today is filled with so much noise. We are overloaded every day with information coming from several sources. Nonetheless, we are somehow capable of selecting our object of attention. This is achieved through our cognitive system's ability to shield the processing of information we have chosen to process, protecting it from other available, but currently irrelevant, information. However, in a given sensory environment, unexpected, abrupt, intense, rare, or simply intrinsically interesting sensory events may occur, which in turn may break through this shield, attracting our attention away from our current focus. Switching attention towards sensory events outside of the current focus is a testament to our cognitive ability for detecting, processing, and evaluating such events. This attention switch has been referred to as the *orienting response*, pertaining to different processing stages (e.g., James 1890, Pavlov 1927, Sokolov 1963b, Sokolov 1963a, Posner 1980, Näätänen et al. 2019). Attention orienting has mainly been investigated within cognitive domains like vision and audition, but it is a crucial mechanism for the linguistic domain too. This is because, as discourse unfolds during spoken communication, interlocutors have to deal with a constantly changing influx of information coming from the speech stream as well as from other channels. The present book therefore explores the role of attention orienting in spoken language by focusing on the relevance of intonation for the orienting response, specifically by investigating the temporal unfolding of its neuro- and psychophysiological underpinnings.

1.1 Topic, research questions, and main findings

Spoken communication necessarily involves *intonation*, the *melody* or the so-called *tune* of our utterances. Intonation conveys meaning over and above the words bearing it. It is encoded primarily through modulations in fundamental frequency (f_0), perceived as pitch variation, but other acoustic properties such as amplitude and intensity (perceived as loudness) or duration (perceived as length), among others, may also play a role. Intonation can differentiate between a number of meanings, and thereby serves multiple functions. For instance, intonation

1 General introduction

can serve to locate fixed attributes of words, like lexical stress in West Germanic languages, it can indicate whether an utterance is a question rather than a statement, or that an utterance is complete or incomplete (for review on the different functions of intonation, see Grice et al. 2023). Crucially for the current work, speakers may also make use of intonation to highlight important parts of an intended message, attempting to draw listeners' attention towards a specific idea they wish to express. Simultaneously, as listeners process their interlocutor's message, their attention is attracted and allocated to information rendered *prominent* through intonation. This is often in the form of a *rise* in pitch, referred to as an *intonational rise*.

Prominence, a notion pertaining to language, is defined as the property of a linguistic unit "standing out" from its neighbouring environment (e.g., Terken & Hermes 2000, Streefkerk 2002, Himmelmann & Primus 2015, von Heusinger & Schumacher 2019, Cangemi & Baumann 2020, Grice & Kügler 2021). Work on linguistic prominence has developed a tripartite definition. Following this definition, prominence is:

- i. a *relational* property, meaning that one can render an element prominent only in comparison to other elements of equal type (see, among others, Cangemi & Baumann 2020),
- ii. a *structural* property, in that the prominent element functions as a structural attractor, licensing more linguistic operations than its competitors (e.g., Himmelmann & Primus 2015), and
- iii. a *dynamic* property, meaning that the prominence status of an entity can shift as time, or discourse, unfolds (see, among others, von Heusinger & Schumacher 2019).

The notion of prominence applies to all linguistic domains, from phonetics and phonology to morphology and syntax, semantics and pragmatics. This means that speakers are able to use a wide variety of linguistic cues to signal prominence, including prosodic and non-prosodic elements (e.g., Baumann & Winter 2018).

Although prominence pertains to the linguistic domain, it features similarities with cognitive notions such as, for example, the notion of *saliency*. More specifically, saliency refers to a sensory event being noticeable within a given sensory environment, by virtue of its physical (also known as *signal-* or *sensory-*

based) characteristics, or its intrinsic (also called *meaning-based*) properties¹ (for review, see Chapter 2). It thus appears that prominence and saliency are much alike, given that they place focus on one entity at the cost of the other entities in an equal set of options. Throughout the remainder of this book, I will show how both prominence and saliency are closely related to the orienting response. For instance, (neuro)cognitive research has shown that stimuli can attract attention because of their physical saliency. More specifically, Chapter 2 shows that in auditory perception, the orienting response is sensitive to the physical properties of the signal, such that the greater the amplitude or pitch rise on a sound, the more enhanced the orienting response towards it (see, among others, Näätänen et al. 1978, Rinne et al. 2005, 2006, Macdonald & Campbell 2011, Liao et al. 2016, Wetzel et al. 2016). In a similar vein, work on prosodic prominence in West Germanic languages has shown that modulations in pitch direction, excursion, and scaling are some of the acoustic dimensions of intonation, also known as *signal-driven* cues, which are indicative of different prominence values. The higher or the more rising the pitch, the greater the prominence value. This makes intonational rises decisive for prominence perception (e.g., Rietveld & Gussenhoven 1985, Gussenhoven & Rietveld 1988, Gussenhoven et al. 1997, Ladd & Morton 1997, Baumann & Röhr 2015, Knight 2008, Niebuhr 2009, Baumann & Winter 2018, Röhr, Baumann & Grice 2022). In particular, Chapter 3 shows that in speech processing, attentional resources are allocated to prosodically prominent constituents, facilitating their processing or their recall performance (e.g., Savino et al. 2020, Ventura et al. 2020, Röhr et al. 2021, Röhr, Savino, et al. 2022). Nonetheless, in addition to signal-driven cues like intonational rises, research has shown that *context-driven* cues, like prior linguistic structure or discourse expectations (or even recent speech experience) have the potential to overwrite the typical acoustic cues that signal prominence, leading to different processes (e.g., Cole et al. 2010, Bishop 2013, Kakouros et al. 2018). Overall, as we will see in Chapter 3, prosodic prominence is expressed in discrete phonological choices, in continuous physical parameters in the phonetic record, and in interactions among acoustic, phonological, and contextual cues (e.g., Baumann & Winter 2018). Therefore, intonation takes on a special role in prominence precisely because the link between signal- and

¹Throughout this book, I make a consistent distinction between *signal-based* and *meaning-based* properties of stimuli. The signal-based properties, as can be inferred from the term used, refer to the physical, acoustic attributes of a respective stimulus. The meaning-based properties refer to the semantics of a respective stimulus, as well as to meaningful aspects licensed by pragmatics and/or contextually created expectations. For this reason, the terms meaning-based, semantic-based, and context-driven are used interchangeably throughout this book.

1 *General introduction*

context-driven cues intrinsically relates to the discrete phonological forms of intonational choices.

Building on previous (neuro)cognitive and linguistic research, the overarching goal of this book is to understand the relevance of intonation for the orienting response in spoken language. This goal can be divided into two central points of interest. Firstly, the attention orienting function of intonational rises is of particular interest for this work, given their inherent association with prominence. I will focus on the following three aspects: the direction of the intonational event (rise vs. fall); the position or status of a pitch rise (and a fall) in relation to prosodic structure (head vs. edge association; pitch accent vs. edge tone); and the role of language- and context-specific expectations for the processing of rises and falls. Accordingly, the following three questions are addressed:

RQ1 [rises]: Do intonational rises attract more attention than falls by virtue of their acoustic prominence?

RQ2 [edges]: Does the phonological status of the rise (and the fall) affect the underlying processes? In other words, do rises at the edges of constituents attract attention to the same extent as accentual rises?

RQ3 [expectations]: Do language- and/or context-specific expectations affect attention orienting?

Secondly, the present work is also concerned with the neurocognitive mechanisms underlying attention orienting in the linguistic domain. Literature on audiovisual attention has established that the orienting response can be involuntary or voluntary. In particular, it has been shown that there is an interplay between the nature of a cue that evokes the orienting response and the attentional processing stage in which this response occurs (for more detail, see Chapter 2). Therefore, the question that this work posits in relation to this second point of interest is the following:

RQ4 [architecture]: To what extent is the neurocognitive architecture of attention orienting similar across the cognitive and linguistic domains?

The main claim of this work is that a plethora of cues can activate the orienting response in spoken language, just as in cognition more generally. Nonetheless, distinct cues are tied to the two core attentional stages. Signal-driven cues appear to be crucial for activating an involuntary attention switch, emerging at a pre-attentive processing stage, while context-driven cues are essential for voluntary

attention orienting, occurring at a conscious processing stage. This claim can be broken down into three parts.

First, I propose that intonational rises indeed take on a special role in activating the orienting response, especially – but not only – pre-attentively. The experimental work reported in this book shows that it is the inherent acoustic properties of rises that make them fundamentally important for involuntary attention-related processes, regardless of whether we are concerned with a simple sine wave, a speech sound in isolation, or meaningful speech. More specifically, the data from the two studies comprising the present work support the previously established idea of an attentional bias towards pitch rises (essentially serving as warning cues), extending it from general cognition to language. Nonetheless, the current work also highlights that, in spoken language, the acoustic signal is not the only source of information. Specifically, the sensory importance of rises is mitigated by the contextual meaning, activating voluntary attention, which in turn is required for updating conscious processing stages and mental representations. Thus, as shown in Chapter 4, when intonational rises are contextually inappropriate, their importance is overwritten at the conscious stage by acoustically less prominent, but contextually appropriate intonational cues, despite that rises take priority over intonational falls at the early pre-attentive stage. However, when all available intonational cues convey the same contextual meaning, then, as Chapter 5 shows, rises take priority over the other available cues, both pre-attentively and consciously.

I further contend that some of the cues pertaining to the activation of attention orienting in spoken language are of primary relevance, while others play only a secondary role. For instance, this work shows that both the *direction* of an intonational contour (rise vs. fall) and the *linguistic meaning* are pivotal for the orienting response. In contrast, I will show that the role of the *phonological status* of the pitch event (accent/edge) is only supplementary. An illustrative example of the primary role of the direction of the intonational contour and the secondary role of its phonological status comes from the current results showing that rising pitch, attributable to both accents and edge tones, is globally more successful in attracting attention than falling pitch, reflected in the neurophysiological and pupillary correlates investigated. Similarly, a contextually appropriate intonational rise,² regardless of its phonological status, can activate both pre-attentive and attentive stages of attention orienting. Yet, a contextually incongruent or inappropriate rise evokes attention only pre-attentively, while attentively its orienting function is “cancelled out” by other, appropriate intonational cues.

²In the present work, appropriate in the list context.

1 General introduction

Finally, I argue that cognition and language appear to share a common pool of mechanisms, crucial for modulating the orienting response both in auditory and speech processing. As Chapter 2 discusses, attention orienting has been linked to both bottom-up (involuntary orienting) and top-down (voluntary orienting) mechanisms, subserved by distinct neural pathways. More importantly, there is an interplay between the mechanisms and low-level sensory cues or high-level cognitive operations: sensory-based cues appear to initiate involuntary attention, feeding bottom-up mechanisms, while meaning-based cues lead to voluntary attention, feeding top-down mechanisms. The current research attests that signal-driven and context-driven prominence cues feed the two core mechanisms related to attention in a similar way. Further, based on the current data, both linguistic prominence and cognitive saliency appear to pertain to similar neural and pupillary underpinnings, providing further evidence for commonalities in the architecture of the cognitive and linguistic domains. Nonetheless, it is important to highlight that, in spoken language, signal-driven and context-driven cues interact with each other in more complex ways than in general cognition, which creates expectations at different layers. The investigation of the incremental processing of linguistic expectations in the current work thus allows for a better understanding of how cues are implemented at different levels (i.e., phonetic, phonological, discourse/context levels) as well as how bottom-up and top-down mechanisms interact with each other during speech comprehension.

1.2 Research rationale

This book investigates the relation between intonation and attention orienting by employing neurophysiological and pupillary data from German listeners. I first elaborate on the rationale behind choosing German as the language under investigation in the present work, and then I turn to the methodologies used.

First off, apart from considerations such as availability of large numbers of participants needed for (neuro)physiological experimentation while conducting the current work in Germany, the choice of German was also motivated by its prosodic characteristics. German is a well-studied West Germanic language with regard to its prosodic system. Among other characteristics, it has lexical stress and postlexical tonal events with specific association and functional properties (e.g., Grice et al. 2005). More specifically, in terms of prosodic typology (e.g., Jun 2014b), in languages like German, there are two basic ways postlexical tonal events can be associated with positions in the prosodic structure. They can either associate as *pitch accents*, docking onto a stressed syllable (head of the word

or foot), or as *edge tones*, pertaining to initial or final edges of smaller or larger constituents. At the same time, pitch accents have been considered to cue prominence, serving a highlighting function, while edge tones have been attributed the function of phonological phrasing. Now, except for the aforementioned functions, intonation also conveys meaning over and above the words or sentences bearing it. This means that different postlexical tonal events convey distinct meanings (without there being a one-to-one mapping), which in turn renders some of them appropriate in certain contexts, while making others inappropriate or surprising. A second reason why German was chosen as the test language here is that prominence-related phenomena have been extensively researched in this language (see Chapter 3). Given that prominence pertains to attention-related processes, previous work on the prosodic marking of prominence and prominence perception in German allows me to build and test specific hypotheses and predictions in the present work. In particular, the attested prominence value of intonational rises offers a useful starting point for further experimental testing of the cognitive and functional contribution of rises to attention orienting, and consequently to prominence. Therefore, the prosodic and prominence properties attributed to the German language offer an excellent test case for investigating the research questions addressed in the present work.

With regard to the methodologies used in the current research, the two experimental studies comprising this book endeavour to answer the research questions introduced in Section 1.1, by investigating the neural (Chapter 4) and pupillary (Chapter 5) underpinnings of attention orienting in spoken language. More specifically, the study reported in Chapter 4 used the method of event-related potentials (ERPs), while the study reported in Chapter 5 employed the method of pupillometry. Both techniques offer a high resolution of temporal information, allowing for the precise measurement of brain and cognitive activity related to attentional processes over time. Previous neurocognitive research has approached the study of attention orienting through measurements of brain activity, focusing on the neurocognitive processes and mechanisms that underlie attention orienting. Specifically, studies using ERPs have revealed specific neural responses, such as the mismatch negativity (MMN) and the subsequent positivity belonging to the P3 family, rigorous correlates of attention orienting. These neural responses have been claimed to reflect the route from pre-attentive to conscious processes (for review, see Näätänen et al. 2019).

More recent psychophysiological research has approached attention orienting using pupillometry. In particular, a growing body of pupillometric research has proposed pupil dilation response (PDR) as a valid index of auditory attention, with similarities to neural MMN/P3 responses (see, among others, Alamia et al.

1 General introduction

2019). For instance, Strauch et al. (2022) have recently shown that the different composites of a PDR can also index distinct processing pathways. Overall, building on these two methodologies used in previous (neuro)cognitive research on attention orienting offers a natural path towards extending the use of correlates which have been shown to be valid indices of the orienting response in general auditory cognition to the study of attention orienting in spoken language. In turn, this extension allows for a more accurate comparison of the function of the orienting response across the two domains.

1.3 Outline and preview of the next chapters

This book is structured as follows:

Chapter 2 sheds light on the concept of attention as an orienting response, elaborating on its origins and the mechanisms proposed to underlie it. After a thorough review of the literature on the involuntary and voluntary nature of the orienting response from early studies up to the modern day, the chapter turns to the function of the auditory processing system. In this context, Chapter 2 discusses the processing stages in which involuntary and voluntary attention emerge, demonstrating the brain indices that underlie these stages. Next, the crucial role of the acoustic properties of the sensory input for attention orienting is discussed, demonstrating the special role of signals involving a rise in amplitude or pitch, the latter being the scope of this book. Finally, this chapter places a special focus on showcasing the methods with which research has approached the study of attention orienting, elaborating on event-related potentials and pupillometry, the two methods employed in the present work.

Chapter 3 offers a condensed introduction of fundamental concepts related to spoken language. Its primary goal is to illustrate the association between intonation and its association to the notion of prominence. After elaborating on the phonological organisation of intonation, the chapter turns to the role of intonation, and especially of rising intonation, in prosodic prominence. To this end, it offers a literature review on the linguistic importance of intonational rises in prominence perception and speech processing. Subsequently, Chapter 3 discusses the interplay between signal-driven and context-driven cues in prominence and speech processing. The chapter closes with a brief note on German list intonation as the linguistic context of the two experimental studies reported in this work.

Chapter 4 reports on the first experimental study. This study examines the processing of intonational rises and falls when presented unexpectedly in a stream of

1.3 Outline and preview of the next chapters

repetitive auditory stimuli. It examines the neurophysiological correlates (ERPs) of attention to these unexpected stimuli through the use of an oddball paradigm. In this paradigm, sequences of repetitive stimuli are occasionally interrupted by a deviant stimulus, allowing for elicitation of the mismatch negativity (MMN). Whereas previous oddball studies on attention towards unexpected sounds involving pitch rises were conducted on non-linguistic stimuli, the present study uses lexical items in German with naturalistic intonation contours as stimuli, potentially giving rise to a list context and the corresponding language-specific expectations. This study mainly relates to RQ1 [rises] and RQ2 [edges], but the results also provide important evidence related to the other key research questions (i.e., RQ3 [expectations] and RQ4 [architecture]).

Chapter 5 introduces the second experimental study. This study addresses the research questions from a different angle. Capitalising on the findings reported in Chapter 4, the study delves further into the attention orienting function of rising edge tones in German, by measuring listeners' pupil dilation response (PDR) in a *changing-state* oddball paradigm. Here, auditory sequences of sequentially ordered (i.e., *seriatim*) ascending numbers (*standards*) are occasionally interspersed with an out-of-the-sequence number (*deviant*). This study mainly relates to RQ1 [rises], RQ2 [edges], and RQ3 [expectations], but certainly, the obtained results inform RQ4 [architecture] as well. Further, this pupillometric study places a special focus on individual variability as a means of better understanding the attention orienting function. In this endeavour, selective attention (in terms of inhibitory ability), processing speed, and working memory capacity (WMC) are selected as proxies for measuring individual cognitive variability. Lastly, the study further explores the role of sequence length in attention orienting. The contribution of both individual cognitive variability and sequence length in attention orienting are particularly important for RQ4 [architecture].

Chapter 6 completes the book by bringing together the findings of the two experimental studies in a general discussion. This final chapter offers a unified view on the cognitive and functional relevance of intonation for attention orienting in the linguistic domain.

2 Attention in the auditory realm

Attention is one of the most prevalent concepts in (neuro)cognition, yet also one of the most diverse and deceptive terms, as researchers have attributed a plethora of phenomena to this concept (for discussion, see Hommel et al. 2019). According to James (1890: 403–404), attention “is the taking possession of the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thoughts. Focalisation, concentration of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed, scatterbrained state...”. Petersen & Posner (2012), revisiting their 1990 paper, propose that attention is an independent system, neuroanatomically different from other cognitive systems, such as incoming stimulus management, decision making, and output production, among others, which, of course, interact with attention and can potentially affect it. As stated in Petersen & Posner (2012), the attention system consists of three neural networks, each one responsible for distinct attentional processes. These are:

- i. the *alerting* network, controlling arousal and vigilance,
- ii. the *orienting* network, regulating sensory processing, and
- iii. the *detection* network, executing target detection, thus also known as the *executive* network.

Although these three networks are responsible for different processes, they depend on each other and cooperate, as the activation of one might lead to the activation of another.

Näätänen (1992), in his book *Attention and Brain Function*, describes some of the principal uses of attention, aiming to arrive at a better understanding of this broad construct. He begins by writing about *selective* attention, which, as he mentions, occurs when we, based on our interests and goals, select to attend one of the many available input sources. To illustrate selective attention, he invokes the well-known cocktail party example, in which many people are talking simultaneously, but listeners can select to attend to the voice of one specific speaker.

2 Attention in the auditory realm

Another example of selective attention is a listener's ability to attend to the linguistic meaning transmitted by a speaker's voice, or the ability to attend to a speaker's voice just as a simple sound event, while disregarding the linguistic message that it is carrying. Thus, given listeners' interests, selective attention can be oriented to a variety of aspects, from simple physical properties, to complex stimuli, as well as to meaningful aspects. Näätänen further describes *divided* attention, i.e., the process of simultaneously attending to different streams of input which are usually dichotically presented: one stream of input is presented to one ear and another stream of input is presented to the other ear. Another phenomenon that has been attributed to the construct of attention is *sustained* attention, which emerges when listeners are asked to detect, through time, infrequent, unpredictable, or weak events. In all these cases described in Näätänen (1992), a form of *active* or *voluntary* attention is involved. Active/voluntary attention is the listeners' ability to choose their object of attention. Thus, voluntary attention shields the processing of information that we have chosen to process from other available but currently irrelevant information. Yet, some events, for example unexpected, abrupt, loud, or rare sounds, outside of the current attentional focus, may break through this shield, involuntarily drawing our attention away from the current focus. This switch of attentional resources has been called *passive attention* (James 1890), *orienting reflex* (Sokolov 1963b, Sokolov 1963a), or *involuntary attention orienting* (e.g., Näätänen et al. 2019), among others. Involuntary attention usually lasts until the stimulus, which initiated the attention switch, is recognised and evaluated, after which attention is drawn back to the initial focus. However, after stimulus evaluation, it can very well be the case that involuntary attention changes to voluntary attention, thus bringing the stimulus to conscious processing. The distinct nature of involuntary and voluntary attention is at the core of this book, and specifically, the neural path from involuntary to voluntary attention orienting is investigated and discussed extensively in Chapter 4.

This book is concerned with the concept of attention as an orienting response, involving different processing stages. It is important to clarify that attention orienting is regulated by all three networks proposed by Petersen & Posner (2012). Over the course of this chapter, we will see that attention orienting is a composite concept whose individual parts pertain to the distinct neural networks. Section 2.1 elaborates on the origins of the attention orienting concept, and presents the two main accounts on the mechanism deemed to underpin attention orienting. Section 2.2 elucidates the difference between the involuntary and voluntary nature of attention shifts through a review from very early to the most recent literature. Subsequently, Section 2.3 is concerned with the auditory processing

system and the stages in which involuntary and voluntary attention emerge, reviewing a model of the brain indices that underlie these processing stages. Section 2.4 continues with a review of studies on the relevance of the acoustic properties of the sensory input for attention orienting, demonstrating the special role of signals involving a rise in amplitude or pitch, the latter being the focus of the present book. Finally, Section 2.5 presents methods with which researchers have studied attention orienting, concentrating on event-related potentials and pupillometry as the two methods that are employed in the current work (see Chapters 4 and 5, respectively).

2.1 What is attention orienting?

In everyday life, we are confronted with an influx of sounds from several different sources. Some of these sounds might be unexpected, rare, or new. Our brains are equipped with mechanisms that can detect unexpected, rare, or new sounds, i.e., sounds that deviate in some property from the current auditory environment. Such deviant sounds may capture attention, prompting shifts of attentional resources in the baseline neural activity. These attentional shifts result from the operation of the orienting network, and can consequently elicit automated motor actions or appropriate adaptive responses if necessary (e.g., Bach et al. 2008). Therefore, in a stimulus-rich environment, sensitivity to auditory changes and evaluation of their importance are essential properties of the neurocognitive system for both survival and communication. For instance, a failure to respond to an important sound change might be life-threatening, and responding to every sound change might exhaust mental resources which are needed for other purposes. In spoken communication, this could entail orienting a listener's attention towards the most important part of the message, which is crucial for effective interpretation and speech planning.

The concept of *attention orienting* has its roots in the seminal work of Ivan Pavlov. In 1927, Pavlov described the concept of an *orienting response* (OR), referring to it as a response with two stages: a reflex-like response towards a change in the current environment, which is followed by a perceptual/conscious processing of this change. Specifically, Pavlov (1927: 12) defines attention orienting as follows: "it is this reflex which brings about the immediate response in man and animals to the slightest changes in the world around them, so that they immediately orientate their appropriate receptor organ in accordance with the perceptible quality in the agent bringing about the change, making a full investigation of it". Pavlov thus proposed an association between the initiation of the OR and an

2 Attention in the auditory realm

early processing stage. During this processing stage, the organism is warned to process an unexpected or new event quickly and efficiently, but the properties of that event are not yet fully processed. At a later processing stage, it is possible to obtain full awareness of this event.

Over the years, many scholars have contributed to the development of Pavlov's OR concept. For example, in the latter half of the 20th century, Posner (1980: 4) defined attention orienting as the "alignment of attention with a source of sensory input or an internal semantic structure stored in memory". Posner further differentiates orienting from *detecting*, a distinct cognitive act, in that the former allows the listener to link responses to the input before it has been consciously processed, while the latter refers to conscious processing of the input. As discussed in Näätänen (1992), views like Pavlov's and Posner's suggest that the OR is evoked by early processes which lead to attention switch, also known as passive attention, or involuntary attention orienting. Section 2.3 elaborates on the different processing stages of attention, and Chapter 4 provides experimental evidence for the link between the pre-attentive processing stage and involuntary attention orienting.

Many studies have further attempted to unravel the mechanism that underpins OR. This work paved the way for the development of many accounts, among them the *novelty detection*, and the *expectancy violation* accounts. The novelty detection account suggests that attention orienting is driven by the detection of a novelty in the current environment, and is rooted in theories which advocate for the role of habituation in OR (e.g., Sokolov 1963b, Sokolov 1963a, Cowan 1998). The OR leading theory by Sokolov (1960) substantially advanced Pavlov's concept. Briefly, Sokolov conceives of orienting as a reflexive response,¹ reflected in physiological activities, such as heart and respiratory rates; electrodermal, vasoconstrictive, pupillary, and muscular reactions; eye, lid, ear, head, and torso movements; or neural changes in the autonomic and/or central nervous system (CNS), among others. He further assumes that the repetition of an originally new event gradually fabricates a memory of its physical properties, resulting in habituation of the OR. This memory has been referred to as the *neuronal model* and is fundamentally important for the novelty detection system: *only* an event that does not match the already-formed neuronal model, i.e., a novel event, has the power to elicit an OR (e.g., Reisberg et al. 1980, Gati & Ben-Shakhar 1990, Elliott & Cowan 2001); otherwise, OR is inhibited.

Despite the simplicity of the novelty detection account in explaining OR activation, this account has received a lot of criticism due to its limitation in address-

¹Although Sokolov's reflex relates to Posner's cognitive act of orienting, it does not allow for a distinction between involuntary and voluntary attention alignment.

2.1 What is attention orienting?

ing and explaining all cases of an OR activation. For instance, an OR activation caused by a familiar but highly relevant (i.e., self-relevant) stimulus, such as one's name in an unattended conversation (e.g., Lynn 1966), cannot be justified by the novelty detection mechanism: a stimulus is rendered familiar precisely because its memory trace (i.e., neuronal model) has already been formed. Thus, according to the novelty detection account, its presentation should not lead to an OR activation, given that *only* a stimulus that does not match the already-formed neuronal model has the potential to elicit an OR. Nor can this mechanism explain why an OR is not elicited by a sequence consisting of completely different but sequentially ordered stimuli such as "1 2 3 4 5 6 7 8...", although in this case a neuronal model for each new digit is missing. Nonetheless, an OR is elicited by the stimulus "5" in a sequence like "1 2 3 5 6 7 8..." (Unger 1964). Lastly, the novelty detection account cannot explain an OR activation by a stimulus which is neither new nor different or relevant, such as for example the stimulus "7" in a sequence like "1 2 3 4 5 6 7 7..." (Velden 1978). In these two last examples, an OR activation appears to arise from interruptions of a sequence generating expectations as to the upcoming item. The "novelty" here, if you like, lies in the unexpected items, which are not the next digits in the sequentially ordered numeric lists.

Given the limitations of the novelty detection mechanism to account for the aforementioned examples, another strand of research suggests that the mechanism underlying the attention orienting response is driven by expectancy violations (e.g., Näätänen 1990, Sussman & Winkler 2001, Vachon et al. 2012). The expectancy violation account thus claims that the sensory processing system develops expectations by detecting regularities (local or global) in the environment. When a deviant sound occurs instead of an anticipated event, it attracts attention, initiating an orienting response. Thus, in this view, the precondition for an OR is the violation of rule-based expectancies and not just the absence of a memory. A large number of experimental studies have provided evidence in favour of the expectancy violation account as being the mechanism that underlies attention orienting (for review, see Paavilainen 2013). Violations of rule-based expectancies have also been referred to as *prediction errors* (e.g., Friston 2010). According to Friston (2010), prediction errors arise because we build an internal model of the world and rely on this model to predict upcoming sensory events. These prediction errors drive top-down expectations towards mental representation updating (Friston 2018). As mentioned in Friston (2018), sensory attention has been associated with noteworthy prediction errors. The expectancy violation basis of attention orienting is further discussed in Section 2.3.

Despite the different mechanisms thought to underlie OR in the two accounts outlined above, many studies tend to conflate the concepts of novelty and de-

2 Attention in the auditory realm

viance (see Tiitinen et al. 1994). This might be because the concept of novelty is not really independent from the concept of deviance, as a new or unfamiliar event typically also violates expectations, for example the stimulus “2” in a sequence such as “1 1 1 1 1 2”. Overall, as Näätänen (1979) summarises, it appears that an OR may actually be initiated by very different aspects of stimuli, such as neuronal mismatches, task relevance, levels of significance, and anticipatory mismatches. Corbetta & Shulman (2002) add to the idea that attention is regulated by different stimulus aspects by stating that both cognitive and sensory cues (as well as their interaction) affect attention. Novelty and unexpectedness are two factors impacting attention that reflect the interaction between cognitive and sensory cues. More specifically, Corbetta & Shulman (2002: 201) posit that “unexpected, novel, salient and potentially dangerous events take high priority in the brain, and are processed at the expense of ongoing behaviour and neural activity”.

A discrepant stimulus, or, in other words, a stimulus that deviates in some way from the recent past, can thus evoke an orienting response. This attention orienting response finds expressions in various (psycho)physiological, neural, motoric, and even behavioural reactions (e.g., Hughes et al. 2007, Liao et al. 2016, Marois et al. 2018, 2019, Näätänen et al. 2019). These reactions have been reported to be indices of the OR. They may occur in different post stimulus stages, thereby reflecting qualitatively distinct types of OR. Going back to Pavlov and to later work by Posner, we can see that both scholars linked the OR with two distinct stages: the pre-attentive stage, related to early processes, and the conscious stage, related to awareness and longer-latency processes. Maltzman (e.g., 1971, 1979) also proposed a distinction between involuntary and voluntary ORs, a distinction rooted in the work by James in 1890. In the next section, I review early theoretical work on the distinction between involuntary and voluntary attention orienting as well as some experimental work that lends credence to this distinction.

2.2 Involuntary and voluntary attention orienting

One of the central topics of the attention literature has been the qualitative difference between the involuntary and voluntary nature of attention shifts. According to Hatfield (1998), speculations about the nature of attentional shifts go back to ancient Greek philosophers and writers of the 4th century who observed that an object or a sound can involuntarily attract attention due to sensory interest. Wright & Ward (2008) mention that Descartes also differentiated between the case of paying voluntary attention to an event of cognitive interest and the case

2.2 Involuntary and voluntary attention orienting

of an event attracting attention involuntarily by causing “admiration”. Descartes further talked about an inverse proportionality between the ability to withhold voluntary attention directed towards an object or a task within the current focus and the intensity of a distracting event outside the current focus, like a loud sound or a pronouncedly novel event: the more intense a distracting event is, the more difficult it is to sustain the processing of the currently attended event.

The literature review of early work on attention reported in Braunschweiger (1899) categorised attentional shifts into two domains: intellectual and sensory. James (1890) discussed attentional processing in these two domains, linking the intellectual domain with voluntary shifts of attention, and the sensory one with involuntary attention shifts. As mentioned in Wright & Ward (2008: 10), for James, voluntary attention entails intentional and conscious preparation of the cognitive processing system, while involuntary attention features an accommodating or adjusting response of sensory organs. James also used the term *passive attention* to describe involuntary shifts, as he reasoned that an unattended stimulus, by virtue of it being intense or sudden, can capture attention automatically. Similarly to Descartes, James wrote about an inverse relation between interest and attention, in that subsiding interest towards an event renders attention more vulnerable to a switch towards unexpected, rare, salient, or novel events.

Introspective and *Gestalt* psychologists also distinguished between involuntary and voluntary attention. Wertheimer (1923), for example, claimed that voluntary attention emerges from voluntary concentration, while involuntary attention arises from stimulus-driven properties. Other scholars described this distinction by suggesting that voluntary attention is directed from one’s self towards an event, while involuntary attention writes in the opposite direction: from an event towards one’s self (e.g., Koffka 1935).

Let us now move to some more recent work. Maltzman reports in his 1979 paper that conditioning of the galvanic skin response (GSR) functions as one proxy of the OR. A basic hypothesis in the classical conditioning of GSR in humans has been the distinction of involuntary and voluntary ORs, resulting from different conditions. The idea is that an involuntary OR is elicited by the initial presentation of an unpredicted stimulus, while a voluntary OR is evoked by task instructions, leading to problem solving or goal-oriented thinking (Maltzman & Mandell 1968). In a series of studies, Maltzman and colleagues (e.g., Maltzman & Raskin 1965, Maltzman 1971), using habituation experiments and measuring GSR as a proxy for OR, provided evidence for the distinction between involuntary and voluntary attention. Briefly, in these studies, two groups of participants were engaged. One group was instructed to respond to a particular word, while the other group did not receive any instructions, serving as controls. The results

2 Attention in the auditory realm

show that both groups responded to the initial presentation of the unpredicted stimulus with the same GSR effect magnitude. However, the group that received the instructions exhibited increased GSRs in relation to the multiple presentations of the stimulus compared to the controls. According to Maltzman (1971), these results suggest that task instructions, leading to self-regulation, modulated the magnitude of the GSR effect in relation to the repeated presentations of the stimulus, although the stimulus had become highly predictable. Based on this finding, Maltzman proposed that GSRs to the initial unpredictable stimulus and GSRs to the subsequent presentations of the increasingly predictable stimulus reflect two different OR types, probably initiated by distinct cortical processes.

As reported in Maltzman (1979), the hypothesis about the distinction of OR into involuntary and voluntary components originates from Pavlov's (1955) signal distinction into first and second systems. The assumption was thus that the first signal controls the involuntary OR because it is a response initiated by some new, unexpected, or intrinsic sensory properties of a stimulus. In turn, the second signal system should regulate the voluntary OR, as the factors that evoke such an OR are instructions, elements of speech, or complex thought processes.

The involuntary and voluntary nature of attention orienting has also been investigated in experimental research using event-related potentials (ERPs).² For instance, Näätänen (1979) reports a series of studies which shed light on the distinction between involuntary and voluntary OR. In this series of studies using selective dichotic listening tasks, participants were presented with streams of frequent signals (standards) including some slightly discrepant stimuli (deviants). The discrepant stimuli were slightly higher in intensity or pitch than standards, and participants were instructed to silently detect and count the deviant stimuli (Näätänen et al. 1978, 1980). In the first two experiments (Näätänen et al. 1978), participants received the same standard and deviant stimuli in both (attended and unattended) ears, while in the 1980 experiment, participants received two different streams per ear, as standard and deviant stimuli were of different properties in the two streams (left ear: 1000 Hz standard and 1150 Hz deviant; right ear: 500 Hz standard and 575 Hz deviant). This series of studies showed that the deviant stimuli always evoked a negative displacement in the signal compared to the standard stimuli, regardless of whether they were presented in unattended or attended conditions. This negative shift was followed by a positive deflection only in the attended condition. On the basis of these results, the authors proposed that involuntary and voluntary ORs are reflected in different ERP components (see Section 2.5). Since the 20th century, a large number of contemporary ERP

²Section 2.5 elaborates on the ERP method.

2.3 Auditory processing stages and their relation to attention

studies have focussed on the neurophysiological underpinnings of attention orienting (for literature review, see Section 2.3, Section 2.4, and Section 2.5) and provided evidence for distinct neurophysiological responses reflecting the activation of involuntary and voluntary OR mechanisms. Specifically, as discussed in the following sections, the sequence of brain events related to attention orienting, depending on the cues that led to their activation, follow a route from pre-attentive to attentive processes, reflecting involuntary and voluntary ORs, respectively.

In summary, the fact that attention can be involuntarily or voluntarily oriented was established in the literature early on. Whereas voluntary (active) attention is initiated by a listener's conscious choice, involuntary (passive) attention is activated directly by a stimulus without the listener's will or control. Stimuli can attract attention because of their sensory or meaning properties. It is evident that there is an interplay between the nature of attention shifts and low-level sensory cues or high-level cognitive operations: sensory cues appear to initiate involuntary attention orienting, while meaningful aspects of stimuli (and thus their derived properties) lead to voluntary attention orienting. It has further been suggested that involuntary and voluntary ORs are controlled by different systems/mechanisms and consequently occur at different processing stages. This proposal foreshadows the contemporary distinction between signal-driven and context-driven attentional processes. Various terms have been used in the literature to refer to this distinction, for example, bottom-up vs. top-down (e.g., Gibson 1966, Gregory 1970), extrinsic vs. intrinsic (e.g., Milner 1974), exogenous vs. endogenous (e.g., Posner 1978), and involuntary vs. voluntary (e.g., Müller & Rabbitt 1989), among other terms. The essence is that the nature of a cue initiates different mechanisms, activating distinct attentional processing stages (pre-attentive or conscious), which results in involuntary or voluntary attention orienting. Table 2.1 summarises some of the related properties of involuntary and voluntary attention orienting based on both the previous literature and the experimental work reported in Chapters 4 and 5.

In the following, I will discuss the different processing stages in the auditory system and their relation to involuntary and voluntary attention.

2.3 Auditory processing stages and their relation to attention

As briefly discussed in Sections 2.1 and 2.2, involuntary and voluntary attention arise at two distinct perceptual stages, the pre-attentive and conscious stage, re-

2 Attention in the auditory realm

Table 2.1: Properties of involuntary and voluntary attention orienting (adapted from Wright & Ward 2008: 27, on visual attention).

Involuntary attention	Voluntary attention
associated with signal-based cues	associated with meaning-based cues
initiated by sensory events	initiated by cognitive events
rapid & transient response	late & gradual response
pre-attentive stage	conscious stage

spectively. One branch of modern cognitive neuroscience has been concerned with the question of which brain processes underlie pre-attentive and conscious perception, resulting in the development of various models of auditory attention and processing. One of the most prevalent models is the “*conscious and unconscious processes in audition*” model by Näätänen et al. (2011).

According to this model, immediately after sound presentation, the auditory system first encodes the sensory features of the sound, then forms its sensory-memory representation (also known as central sound representation, CSR; Näätänen & Winkler 1999) by integrating feature and temporal information, and finally shifts to a slowly decaying phase that represents the sensory-memory trace of that particular sound. Based on the sensory-memory traces of the preceding sound sequences, the auditory system detects regularities between successive sound events, extrapolates a pattern, and predicts the properties of the prospective sound events. When these anticipatory predictions are not attested, in other words when they are violated in that the properties of the upcoming sound event deviate from what is expected, attention-calling mechanisms in relation to this unexpected event are activated (for details on the model of central auditory processing, see Näätänen et al. 2011, 2019).

It has been shown that the first attention-calling mechanism that is activated in the auditory cortex after the presentation of an unexpected auditory event is the automatic change-detection signal, which evokes a neurophysiological response called mismatch negativity (MMN) (e.g., Näätänen 1992). MMN is defined as a regularity-violation response elicited by any perceptible change in auditory stimulation, even during unattended listening (for reviews, see Näätänen et al. 2007, 2019). MMN elicitation in turn alerts the executive mechanisms, which might result in a physiological or behavioural response. MMN thus represents an automatic, pre-attentive response, activating an involuntary attention switch towards an unexpected auditory change (e.g., Näätänen 1990, 1992, 2001, Näätänen et al. 2019).

2.4 *The role of rises in attention orienting*

After MMN elicitation, and thus an involuntary switch of attentional resources towards the unexpected auditory deviance, this deviance can be brought into awareness. Conscious perception can be initiated either by a strong attention-calling process (Näätänen 1992) or through the formation of an attentional trace. The formation of the attentional trace depends on the conscious, voluntary maintenance of the deviance's features by the attentional control mechanisms. As discussed by Näätänen et al. (2019), when the attentional trace is voluntarily maintained (this voluntary maintenance of the attentional trace is also called *rehearsal*), the deviance (and its properties) is, first, available to top-down processes, and second, has the potential to reach long-term memory, with both cases leading to its conscious perception and semantic activation. Studies have shown that the route from pre-attentive to conscious perception of an unexpected auditory change or deviance is reflected through the MMN-P3 complex. This complex consists of the MMN activation, followed by a positive deflection. This positive deflection has been identified within the P3 family. The P3 response has been reported to reflect conscious processes of novel or salient events (e.g., Donchin 1981, Polich 1986, Duncan et al. 2009).

Therefore, one could argue that after the presentation of an unexpected deviance in the current auditory environment, the MMN plays a fundamental role in the sequence of brain processes that underlie attention. First, the MMN generator automatically shifts involuntary attention towards the unexpected auditory deviance, and subsequently, depending on the attentional trace of this deviance, activates the mechanisms of conscious attentional processes. In other words, attention is voluntarily oriented towards the deviance, reflected in the presence of a P3 response, by which this change is brought into awareness. In Section 2.5, I expand on the properties of these two neurophysiological correlates, as they are of particular interest for the research reported in Chapter 4. Prior to that, Section 2.4 focuses on the relevance of the physical properties of the sensory input for attention orienting. In particular, the aim of the following section is to demonstrate the special role of signals involving a rise, for example in amplitude or pitch, for attention orienting, with the latter being the focus of the experimental Chapters 4 and 5.

2.4 The role of rises in attention orienting

As we have seen in Section 2.2, the idea of an inverse relation between a listener's capacity to voluntarily sustain attention towards an event and the physical intensity of a distraction has been discussed in the literature for a very long time. One

2 Attention in the auditory realm

branch of modern neurocognitive empirical research has also been concerned with the question of how attention orienting towards unexpected sound events is conditioned by different auditory cues. The relevant studies manipulate the sound change direction by modifying a variety of acoustic features such as fundamental frequency (f_0 ; perceived as pitch), duration (perceived as length), and intensity (perceived as relative loudness), and have reported an attentional bias towards unexpected sounds with rising as opposed to falling acoustic properties, indexed by a greater MMN amplitude or elicitation of an MMN-P3 complex (for review, see Näätänen et al. 2019). The last point deserves special attention, as the contribution of pitch rises to attention orienting is at the core of this book. Therefore, in what follows, I review some of the neuroscientific studies which have contributed empirical evidence to this idea in the domain of cognition.

In one of the earliest studies using ERPs, Näätänen et al. (1978) presented participants with sequences of frequent sounds, occasionally interrupted by sounds deviating in loudness (experiment 1) or pitch (experiment 2). Specifically, deviants were of higher loudness and pitch compared to the frequent sounds. In both experiments, deviant sounds (with higher loudness and pitch), elicited a negative deflection in the signal, similar to the MMN response which was later discovered. Näätänen et al. (1989) investigated ERP responses to task-irrelevant infrequent tone pips occurring in a stream of frequent tone pips. In particular, the infrequent sound events were characterised by increasing or decreasing intensity compared to the frequent sounds. The analysis was focused only on the decreasing deviants, reporting elicitation of an MMN response, with larger amplitude and shorter latency, as the deviants' intensity decreased. Further, in an MMN study, Alain et al. (1994) presented participants with sequences of two tones, say A and B, alternating repetitively, e.g., "A B A B A B...". This repetitive alternation was interrupted from time to time by the same tone being presented twice, e.g., "A B A B A **A** B...". The tones differed in frequency by one, six, or twelve semi-tones, but all the remaining properties (intensity, duration, etc.) were kept the same. Participants were instructed to ignore the auditory stimuli and their task was to read a book of their choice. The authors report that all tones deviating from the pattern elicited an MMN response. Nevertheless, MMN amplitude was larger when the difference between the two tones was an increase in frequency.

Subsequently, Rinne et al. (2005) used functional magnetic resonance imaging (fMRI) to investigate the processing of infrequent large, medium, and small duration decreases in a stream of repetitive sound duration during unattended listening. The authors' prediction was that brain activation, as reflected in haemodynamic responses, would increase with a larger decrease in duration. This prediction was based on previous fMRI studies on auditory detection of pitch changes

2.4 The role of rises in attention orienting

in which the magnitude of the brain response was positively correlated with the magnitude of the sound change, although these studies explored sound changes with pitch rises (e.g., Doeller et al. 2003). In contrast to their expectations, Rinne et al. (2005) find that brain activation tended to be greater with small and medium duration decreases than with large ones. The authors tested the same contrasts in a separate ERP session, which also showed no differences in MMN amplitude among the different magnitudes of the sound decreases. Next, Rinne et al. (2006) used ERP measures to study the neural responses to intensity changes during unattended listening. Specifically, participants were presented with a repetitive stream of sine waves with a standard intensity of 60 dB interspersed with infrequent intensity falls and rises. The results show that both infrequent intensity falls and rises evoked an MMN, but also that rising intensity elicited an additional P3 response. Likewise, Bach et al. (2008), using fMRI, measured brain activity in relation to infrequent rising and falling sound intensity. Their results show that rising sound intensity (as opposed to falling sound intensity) led to activation of the right amygdala and left temporal areas of the brain, accelerating the activity response time.

In an ERP study, Macdonald & Campbell (2011) tested whether a *psychological* rise in intensity would be processed differently from a *psychological* fall in intensity, in a similar way to physical intensity rises and falls. Before I move to the results of this study, let me elaborate on what the authors mean by referring to *psychological* rising and falling intensity. The stimuli of this study consisted of sequences of sine waves following a rule-based alternating low-high intensity pattern, as in the following sequence: L H L H L H. Now, this rule-based pattern was occasionally interrupted either by a repetition of the high-intensity sound (e.g., L H L H **H** H L H), or by a repetition of the low-intensity sound (e.g., L H L H L **L** L H). According to the authors, the repetition of the high-intensity stimulus in the first example sequence created the percept of a relative, psychological, intensity rise because the rule would have predicted the low-intensity sound. Likewise, the repetition of the low-intensity stimulus in the second example sequence created the percept of a relative, psychological, fall because the rule would have predicted the high-intensity sound. The results of this study show that MMN peak latency and amplitude to psychological rising sound intensity were significantly earlier and larger, respectively, compared to psychological falling sound intensity. The authors conclude that a deviant representing a rise in intensity, compared to what the auditory past would have predicted, is processed in a different way compared to a deviant representing an intensity fall, even when this rising/falling property is a controlled construct and not a physical attribute.

2 Attention in the auditory realm

Chobert et al. (2012) investigated the pre-attentive auditory processing in children with dyslexia compared to matched controls. The authors used syllables with Consonant-Vowel (CV) structure as stimuli and tested ERP responses to f_0 and vowel duration deviants. Specifically, f_0 deviants were of higher pitch, while duration deviants were of shorter length, compared to standards. The authors report that all deviants evoked an MMN response, in both groups. Specifically, for the f_0 deviants, the higher the pitch, the larger the MMN amplitude. However, for duration deviants the magnitude of duration decrease did not affect the magnitude of the MMN response.

In a more recent study, Hsu et al. (2015) explored the sensitivity of pitch change direction in relation to speech. It is important to note that, although some of Hsu et al.'s stimuli were produced with a human voice, they were not designed to convey any linguistic meaning. Specifically, the authors tested pitch changes produced on the phoneme [a] with the Mandarin level tone, starting from habitual pitch and increasing f_0 for as long as modal phonation was possible. For a second set of stimuli, they synthetically elevated the pitch above the speaker's normal range. Pitch changes in non-speech stimuli (i.e., pure tones) at the same frequencies were also included in the study. Using an oddball paradigm, Hsu et al. investigated whether unexpected small and large rises in spoken pitch attract listeners' attention to a greater extent than unexpected small and large pitch falls. In addition, they asked whether brain responses to spoken pitch rises are different from similar rising pitch changes in pure tones. They found that whereas MMN to changes at normal and synthetically elevated spoken pitch height did not differ as a function of pitch direction (falling vs. rising), it did differ as a function of the size of the change (small vs. large), as large pitch changes evoked a greater MMN. MMN to pure tones equivalent to the speaker's normal pitch height was evoked only by large falling changes in pitch, while at elevated levels, it was evoked by small and large pitch rises, and by large pitch falls. Additionally, P3 was sensitive to the direction of the change, since only rising pitch changes (both small and large) at a normal spoken pitch height evoked a P3. With pure tones, large pitch changes at an elevated pitch height also gave rise to a P3 which, however, was not sensitive to pitch direction. Based on the above results and, specifically, the P3 sensitivity to rising pitch changes in speech, the authors suggest that sudden pitch rises in speech demand more attentional resources than sudden falls, since their presence activates additional conscious processing mechanisms.

Considering the results reported in the studies summarised above, it is evident that the saliency of the physical properties of the acoustic signal is essential for attracting attention, highlighting the orienting function of signals involving a

2.5 Approaches to studying attention orienting

rise in intensity, duration, or pitch. Bach et al. (2008) mention that rising acoustic signals prompt auditory looming effects, indicating that the sound source is approaching. Some of the aforementioned studies further show that falling acoustic signals may also attract attention, but it appears that they are experienced mostly as fading (indicating a receding sound source). Falling signals are thus processed differently, activating attentional resources in a way that is distinct from the processing of rising signals (e.g., Rinne et al. 2005, Macdonald & Campbell 2011). Rises can then essentially be seen as intrinsic warning cues, activating attentional resources, which in turn elicit a series of neural, psychological, or physical reflexive responses (Bach et al. 2008). Chapters 4 and 5 revisit the idea of an attentional bias towards rising sounds, focusing on pitch rises (comparing them to pitch falls), and extend this idea from general auditory cognition to spoken language by investigating the neurophysiological and pupillary underpinnings of attention orienting in German intonation.

2.5 Approaches to studying attention orienting

The scope of this section is the review of methods employed by researchers to study attention orienting. Special focus is placed on two techniques, event-related potentials and pupillometry, as these techniques are highly relevant for the experimental work presented in Chapters 4 and 5.

2.5.1 Behavioural measures

As discussed above, the orienting response has behavioural, physiological, psychological, and neural manifestations. Within the cognitive-behavioural tradition, a strand of research has been concerned with auditory attention capture and its behavioural correlates (e.g., Cowan 1998, Näätänen et al. 2001). This body of research investigates attention orienting towards an auditory distraction using tasks like visual serial recall (e.g., Hughes & Jones 2003, 2005), categorisation of visual targets (e.g., Escera et al. 1998), and detection of auditory targets in dichotic listening (e.g., Näätänen et al. 1993), among other tasks. In these experimental set-ups, participants are simultaneously presented with sound sequences including some deviances, but are explicitly instructed to ignore them, as they are irrelevant for the task they have to perform. Such auditory deviances draw attention away from the ongoing task, inducing disengagement. Attention orienting towards the deviances is expressed through the disruption of the ongoing activity, leading to longer processing or response time, and/or deterioration in task performance, even if the task is in a different modality than the deviant

2 Attention in the auditory realm

sound. Some studies have also shown that physiological responses to auditory deviances and performance deficiencies often co-occur, highlighting an association between them (e.g., Näätänen et al. 2001).

2.5.2 Traditional psychophysiological methods

A vast array of methods investigating OR has been employed in traditional psychophysiology, including electrodermal (e.g., Maltzman 1979), electromyographic (e.g., Dimberg 1990), vascular (e.g., Unger 1964), and cardiac measures (e.g., Graham & Clifton 1966), among others. In these experimental set-ups, an auditory deviant stimulus activates a cognitive process which in turn evokes a physiological response. In other words, indices of changes in the body are tracked which are evoked by certain events in the information stream. Most of these psychophysiological measures exhibit long latencies (i.e., emerge more than 2 seconds after stimulus onset) relative to the rapidity of the corresponding cognitive processes, and they therefore provide a sum response at the end of the physiological change. They serve thus as indirect measures of the activated processes, complementing more direct measures. As a result, these measures do not allow for an analytical investigation of the different processing stages associated with the orienting response through the course of the observed physiological changes, as they can lead to confounds in the interpretations of the orienting response (for a more elaborate discussion, see Näätänen 1992). Nevertheless, it is crucial to clarify that most of these measures may be sluggish in relation to the cognitive processes, but they are rapid in relation to pre-attentive/activational and emotional processes (Näätänen 1992). In this sense, whereas these measures do not allow for a distinction between involuntary and voluntary OR mechanisms, they do provide correlates of passive attention, and thus the involuntary component of OR. This is one of the reasons why the relevant measures were fundamental for early research on the concept of the OR, and why these methods are still of great importance for OR and for other fields of psychophysiology.

2.5.3 Brain measurements: The event-related potentials technique

Neurocognitive studies have approached attention orienting through measurements of brain activity. This line of research has been particularly concerned with the neurocognitive processes and mechanisms that underlie attention orienting. Nowadays, many methods are available for measuring brain activity, such as electroencephalography (EEG), magnetoencephalography (MEG), positron emission tomography (PET), single photon emission computed tomography

(SPECT), thermoencephalography, and functional magnetic resonance imaging (fMRI), among others. EEG measures the electrical activity in the brain recorded simultaneously from many sites on the scalp, MEG measures neural activity in specific brain areas via a magnetic field, and the remaining techniques measure brain activity via changes in the blood flow or metabolic rate (for review of the different brain measurements, see Näätänen 1992). Whilst the EEG method offers a high temporal resolution of electrical changes in the brain, all other measures provide high spatial resolutions of brain activity in locating the generating area. Given that this book is concerned with the temporal unfolding of the neurophysiological underpinnings of attention orienting, I focus on the EEG method, which offers a high resolution of temporal information. This technique is implemented in the experimental investigation reported in Chapter 4.

Traditional psychophysiology has also used EEG α -blocking to measure the alpha-rhythmic activity of the brain in response to auditory stimuli, intending to reveal possible task-specific patterns in this activity (see Näätänen 1992). However, borrowing Luck's (2014: 4) words, "EEG is a very coarse measure of brain activity", as the raw signal consists of merged neural activities arising from many different sources, making it almost impossible to identify specific neurocognitive responses, for example those related to attention orienting. Thus, the EEG method was not considered effective in traditional psychophysiology (for a critical review, see Beaumont 1983). However, the neural responses related to specific sensory or cognitive events, which co-exist in the EEG signal along with neural activity emerging from other sources, can be identified within the overall signal using the EEG-based method of *event-related potentials* (ERPs; Luck 2014).

The ERP method is a simple averaging technique, and, as inferred by its name, refers to electrical potentials, namely, small changes in brain activity, which are temporally related to specific sensory or cognitive events (e.g., Luck 2014, Bornkessel-Schlesewsky & Schumacher 2016). ERPs are recorded via electrodes mounted in an elastic EEG cap and placed on the scalp. As previously discussed, ERPs are extracted from the neural activity by averaging across multiple instances of the same stimulus, separately for each electrode site. Specifically, averaging is applied to a specific time window, starting at stimulus onset and lasting up to a time point defined by the researcher. This produces the averaged ERP waves consisting of positive and negative electrical deflections over time, which are called *peaks*, *waves*, or *components* (Luck 2014). Figure 2.1 presents a graphic illustration of the ERP technique.

ERP components are characterised in terms of the following attributes:

- i. *polarity*, a deflection's positive or negative configuration,
- ii. *latency*, a deflection's peak point from stimulus onset,

2 Attention in the auditory realm

- iii. *topography*, a deflection's scalp distribution over electrode sites, and
- iv. *amplitude*, a deflection's area under the curve, signifying its magnitude.

ERP components are mainly classified into three categories: exogenous, endogenous, and motor components (e.g., Donchin et al. 1978, Luck 2014), although the boundaries between these categories are often blurred. The exogenous ERP

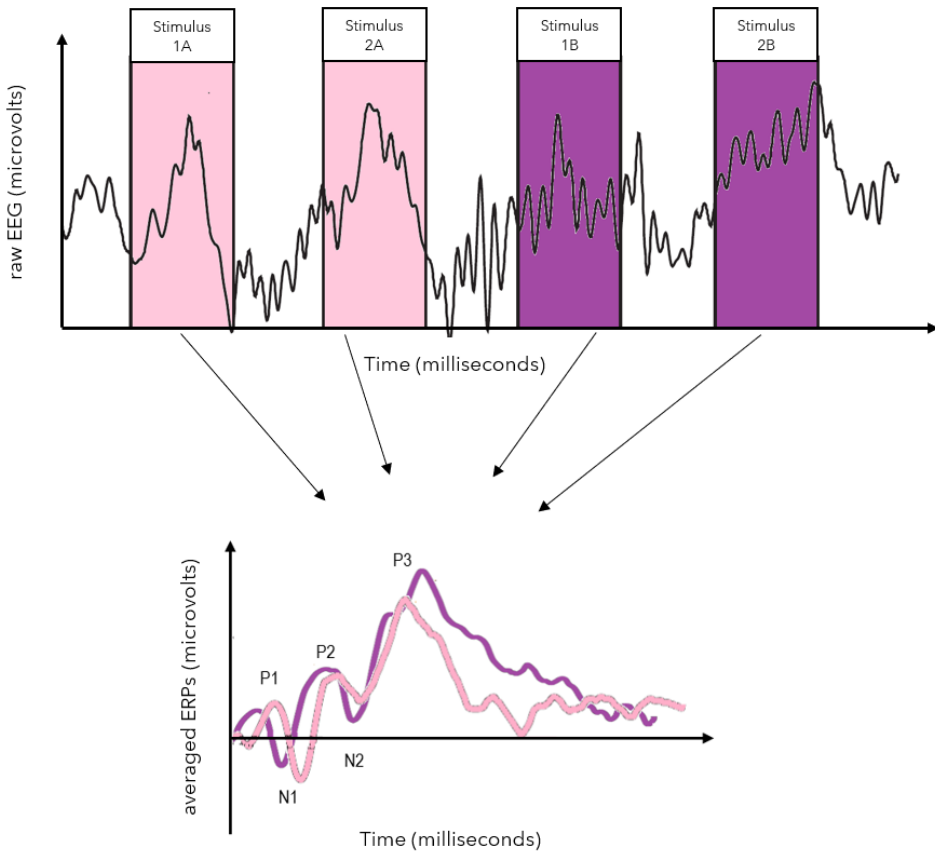


Figure 2.1: The continuous raw EEG signal (in microvolts, positivity upward) is shown at the top of the figure over a period of time. The pink rectangles illustrate EEG segments corresponding to presentations of stimulus A, and the purple rectangles depict EEG segments corresponding to presentations of stimulus B. The individual EEG segments are extracted and averaged as a function of stimulus, in order to obtain the two averaged ERP signals shown at the bottom of the figure. (Adapted from Luck 2014: 7).

2.5 Approaches to studying attention orienting

components are defined by the external, i.e., sensory, attributes of a stimulus. The endogenous ERP components are triggered by cognitive processes. Lastly, the motor ERP components are tied to a given motor response.

Further, based on their classification and attributes, ERP components receive a specific label. Traditional ERP labels consist of a letter, *P* or *N*, indicating positive or negative polarity of the component, and a number, reflecting either the peak's position in the signal (i.e., if it is first, second, third peak etc.), or the peak latency in milliseconds.³ For instance, an ERP component with positive deflection that occurs 300 ms after stimulus onset, which also happens to be the third positive peak in the signal, is labelled either *P3* or *P300* (see Figure 2.1 for an illustration). ERP components may also be linked to particular functions, and are then named after their functional relevance. One example is the *error-related negativity*, a component which is elicited by an error in the stimulus stream (for more on the naming conventions of ERP components, see Luck 2014).

The ERP technique is widely used in cognitive neuroscience, sensory processing, and language processing (of both spoken and written stimuli), among other domains. It is a powerful tool compared to behavioural or traditional psychophysiological measures, as it provides a high resolution of temporal information, in turn making it possible to continuously measure brain activity in response to a stimulus with millisecond accuracy (e.g., Näätänen 1992, Luck 2014). This allows for an analytical investigation of the neurocognitive processes before, after, or during a specific stimulus event, in other words, as time unfolds. Further, as discussed by other scholars, ERPs offer a measure of online processing without demanding a behavioural response, i.e., during unattended stimulation (e.g., Näätänen 1992, Luck 2014). This aspect is particularly useful for the study of attention orienting, as it allows for the investigation of the path from pre-attentive to attentive processes during unattended stimulation.

A large body of ERP research on automatic sensory processing has reported a number of early and late exogenous ERP components related to audition, for example, the Auditory Brainstem Responses (ABRs), the Middle Latency Responses (MLRs), and the Late Transient Responses (N1, P2, N2, P2). All these exogenous auditory ERP components, which range from 10 ms to 150 ms post stimulus, are related to the physical characteristics of the stimuli (for review, see Näätänen 1992). In other words, they are signal-evoked potentials. Moreover, many ERP studies on attention orienting have been instrumental in elucidating the association between specific ERP components and attentional processes, namely the mismatch negativity (MMN) and the MMN-P3 complex, as already introduced in

³Usually, when the number is greater than 5, it refers to peak latency (Luck 2014).

2 Attention in the auditory realm

Section 2.3. The MMN and P3 components (in contrast to the exogenous auditory ERPs) are endogenous, as they are generated by a regularity-change-detection system, reflecting cognitive processing. In what follows, I elaborate on these two ERP components, which are central to the study reported in Chapter 4.

2.5.4 ERP correlates of attention orienting: MMN and P3

The mismatch negativity, known as MMN, is a neurophysiological response recorded in the EEG,⁴ and discovered by Näätänen et al. (1978) in the late 1970s. Näätänen et al. (2019: 1) describe MMN as “a window to central auditory processing” which enables us to better understand “the brain processes forming the biological substrate of central auditory processing and the different forms of auditory memory, as well as the attentional processes controlling the access of sensory input to conscious perception and higher forms of memory”. As mentioned in Section 2.3, MMN is elicited by any perceptible change in the auditory stream. It is well established that MMN elicitation is independent from attentive processing (for reviews on MMN in passive recordings, sleep, patients in coma, and anaesthetised humans & animals, see Näätänen et al. 2007, 2019). Accordingly, it has received the label *automatic regularity-violation response*.

MMN can be recorded in different stimulus paradigms, depending on the aspects of the central auditory system under study. One of the most frequently used paradigms is the classic *steady-state* oddball paradigm in passive recordings (e.g., Näätänen et al. 2019), which is also employed in the experimental study presented in Chapter 4. In this paradigm, listeners are presented auditorily with sequences of repetitive sounds (called standards) which are occasionally interspersed with a sound that deviates in some property from the steady-state stream (called deviant) while they are watching a film with no sound, or reading a book of their choice.⁵ Another commonly used paradigm is the *changing-state* oddball paradigm, in which listeners are presented with sequences of changing auditory stimuli leading to a pattern (standards), occasionally replaced with a stimulus that disrupts the pattern (deviant). The changing-state oddball paradigm is used in the experimental study presented in Chapter 5. Finally, some studies have utilised the *multi-feature* oddball paradigm, in which every second stimulus is a standard, while the rest are deviants of different types (for reviews of the different MMN paradigms, see Paavilainen 2013, Näätänen et al. 2019). In all oddball

⁴MMN can also be recorded in MEG, called MMNm. In addition to the auditory modality, MMN can be recorded in the visual (vMMN), somatosensory (sMMN), and olfactory (oMMN) modalities (for more, see Näätänen et al. 2019).

⁵Oddball paradigms have been also used in active recordings, or dichotic listening tasks.

2.5 Approaches to studying attention orienting

cases, the deviant stimuli evoke an MMN activity in relation to their standard stimulation (see Section 2.3 for more on the MMN generation process).

Figure 2.2 depicts an MMN to small frequency deviants in a stream of repetitive 1000 Hz tone stimuli, during unattended processing. The MMN is a negative potential which is traditionally obtained as a difference wave by subtracting the ERPs to standard from those to deviant stimuli (i.e., deviant ERPs – standard ERPs; Näätänen et al. 2019). MMN has a fronto-central maximum and a peak latency between 150 and 250 ms after the stimulus onset. Nevertheless, there is considerable variability in the definition of this time window, as peak latency has usually been defined on the basis of difference waves.⁶ Moreover, latency may vary with the degree of deviation (e.g., Emmendorfer et al. 2020). There are other transient brain responses that indicate some kind of change or deviance detection (e.g., P3a/b, N400, P600), although their elicitation requires some level of saliency or directed attention. These can, for instance, be triggered by task demands (Sussman et al. 2014).

MMN has been described as a correlate of involuntary attention switch towards deviant sounds in the current auditory environment, as its elicitation reflects the automatic activation of the first attention-calling mechanism in the auditory cortex. As mentioned above, MMN is elicited rapidly after stimulus onset, revealing attentional processes that emerge at the pre-attentive stage. As outlined in Section 2.3, MMN generation can trigger an “attention leak” towards the unattended deviant, bringing it into awareness. In other words, MMN can initiate subsequent attentional processes at the attentive stage after activating an involuntary attention switch. This is usually reflected in the activation of additional ERP components such as the MMN-P3 complex. The MMN-P3 complex consists of MMN activation accompanied by a positive deflection within the P3 family (for review, see Näätänen et al. 2019). The P3 is a broad, positive-going component with peak latency around 300 ms (or more) after stimulus onset (P3 latency depends on the complexity of processing – the more complex the processing, the longer the latency – with ranges from approximately 250 to 1000 ms; see Duncan et al. 2009). P3 has been divided into two subcomponents, P3a and P3b (for review, see Polich 2007). P3a has a fronto-central scalp distribution with shorter

⁶As Luck (2014: 349) mentions “any amplitude or latency measurement procedure that involves finding a peak is non-linear”. Therefore, measuring peak latency before computing difference waves (or averaging waveforms together) leads to different results than performing the same measurement after computing difference waves. Further, latency differences may arise from amplitude changes. Although difference waves are a great tool for isolating specific components from the parent waveform, latency estimates can be “distorted by changes in the amplitude of temporally overlapping components” (Luck 2014: 57).

2 Attention in the auditory realm

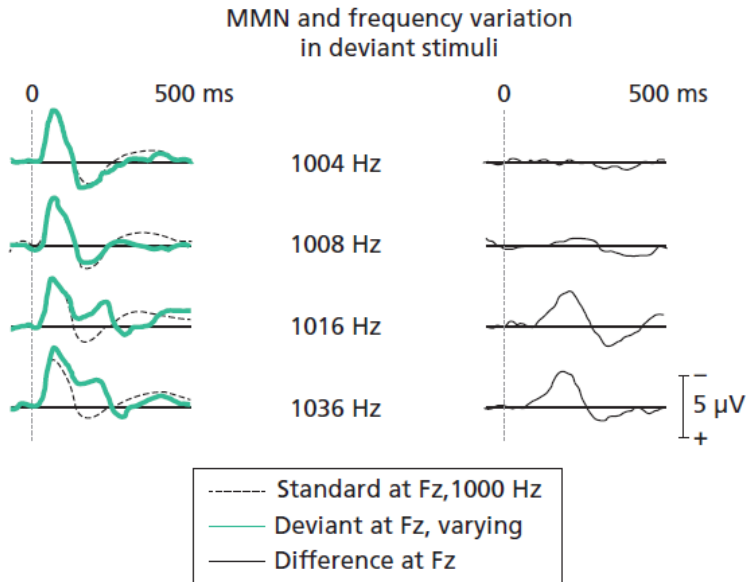


Figure 2.2: ERPs to standard and deviant stimuli are depicted on the left-hand side. Difference waves between ERP to deviants and ERP to standards, indicating the MMN component, are illustrated on the right-hand side. Both ERP and difference waves are shown at the Fz^a electrode site (figure taken from Näätänen et al. 2019: 2).

^aThe Fz electrode site refers to the placement of an electrode at the frontal (F) brain region on the midline (z). See Chapter 4 for more information on electrode placement according to the 10–20 system.

and sharper peak latency, and has been related to attentive processes. P3b has a posterior scalp distribution with longer peak latency, and has been related to memory during discourse updating (for reviews, see Polich 2007, Duncan et al. 2009). The current work is specifically concerned with the P3a, as it has been reported to reflect conscious attentional-related processes of novel, less probable, or salient events (in terms of intrinsic significance or self-relevance and not physical properties; e.g., Donchin 1981, Polich 2007, Duncan et al. 2009). Thus, MMN and P3 responses have been identified as correlates of attention orienting, reflecting the route from pre-attentive to conscious processing of an unexpected auditory change or deviance (e.g., Escera et al. 1998, Näätänen 2001, Hsu et al. 2023).

2.5 Approaches to studying attention orienting

Except for a correlate of attention orienting, clinical studies have also outlined MMN as an indicator of general brain plasticity (e.g., Kreegipuu & Näätänen 2009), with wide clinical use. For example, MMN deficiency has been found to reflect various neuropsychiatric, neurological, and neurodevelopmental cognitive and functional profiles (for an elaborate review of MMN applications, see Näätänen et al. 2019). The MMN component has been further used to study linguistic processes, from simple phoneme discrimination (for review of studies, see Näätänen et al. 2019) to high-order processes such as syntax (e.g., Hasting et al. 2007), revealing MMN as an index of language-specific detection, reflecting long-term traces for linguistically meaningful elements in the brain. Likewise, P3 has been associated with attentional processes during online comprehension of spoken language (e.g., Röhr et al. 2021). Certainly, there are various ERP components relevant for language processing (e.g., P200, N400, Late Positivity, among others), which, however, are beyond the scope of this book, and thus will not be discussed here (for a brief review on language-related ERP components, see Luck 2014). MMN and P3 responses in relation to speech processing will be further reviewed in Section 3.4 of Chapter 3. In Chapter 4, MMN/P3 responses are revisited and explored as neurophysiological underpinnings of attention orienting in spoken language.

2.5.5 Indexing attention orienting using pupillometry

Another psychophysiological method that has been employed to study attention orienting is *pupillometry*. Pupillometry is a widely used method, applied in various scientific domains, measuring changes in pupil size as a proxy for distinct neurocognitive processes (e.g., Amiez & Procyk 2019).

The black aperture at the centre of the iris (the coloured part of the eye) is called the pupil. Light passes through the pupil, with the iris controlling the amount of light that passes through. From there, light hits the retina, and is converted into an electrical signal sent to the brain via the optic nerve. Luminance changes evoke changes in pupil size. Specifically, high levels of light result in pupil contraction, while low levels of light lead to pupil dilation (e.g., Strauch et al. 2022). The pupil light reflex is mediated by two antagonistic networks: pupil contraction is controlled by the parasympathetic system, and pupil dilation is regulated by the sympathetic system (for more on the pupil control pathways, see Amiez & Procyk 2019, Strauch et al. 2022).

As Johansson & Balkenius (2018) point out, it has been established that whilst large changes in pupil size emerge as a result of luminance changes, pupillary responses also occur as a function of distinct brain processes. Specifically, the

2 Attention in the auditory realm

pupil dilation response (PDR) has mainly been correlated with processes such as cognitive effort (in the sense of deliberate allocation of mental resources during a cognitive task), emotional processing, heightened attention, expectancy violation, and memory consolidation, among others (for review, see Winn et al. 2018). Many of these studies have measured pupillary responses in relation to auditory stimuli (for review, see Zekveld et al. 2018). In particular, a growing number of pupillometric studies have employed pupillometry to investigate auditory distractions and proposed the PDR as a valid psychophysiological index of auditory attention orienting, similar to the neural MMN/P3 responses (e.g., Nieuwenhuis et al. 2011, Wang & Munoz 2015, Johansson & Balkenius 2018, Marois et al. 2018, Alamia et al. 2019).

Strauch et al. (2022) suggest pupillometry as a valuable psychophysiological tool for investigating the differential neural networks of the attention system as proposed by Petersen & Posner (see beginning of this chapter). The authors further propose a classification of pupillary responses as a function of the factor that led to their elicitation, and rank these factors as:

- i. *low-level*, including luminance changes and focal distance,
- ii. *intermediate-level*, involving alerting and orienting, and
- iii. *high-level* factors, encompassing an interaction between conscious attention and sensory processing.

Each of these factors activates different neural circuits, which in turn elicit distinct types of pupillary responses. In particular, as already mentioned above, pupillary responses triggered by low-level factors are stimulated by the parasympathetic and sympathetic systems. Pupillary responses evoked by intermediate-level factors are mediated by two distinct pathways in the subcortical areas: arousal-related PDRs result from the locus coeruleus system (LC), and orienting-related PDRs are controlled by the superior colliculus system (SC) (e.g., Wang & Munoz 2015, Joshi & Gold 2020). Finally, pupillary responses to high-level factors are regulated by the executive network located in the frontoparietal areas.

The pupil is characterised by two states: a steady-state with a baseline size, and a transient-change state indicating temporary deviations from the baseline, expressed in sudden dilations or contractions (Joshi & Gold 2020). Low-level factors lead to size changes in the pupil steady-state. Orienting factors trigger transient-changes in the baseline pupil size. Finally, alerting and executive factors can result in both steady-state and transient pupil size changes depending on whether the effect is sustained or temporary.

2.5 Approaches to studying attention orienting

Orienting-related PDRs are usually momentary, meaning they have short latencies: their onset typically occurs between 200 and 500 ms post stimulus, peaking approximately 1 second later, and ending rapidly just after stimulus completion (for review on pupillary latencies, see Beatty 1982). Further, orienting-related PDRs are sensitive to the saliency of the physical properties characterising auditory deviances. Specifically, the greater the acoustic saliency of a deviant (i.e., rising sounds), the larger the PDR amplitude (e.g., Liao et al. 2016, Wetzel et al. 2016, Marois et al. 2018, Strauch et al. 2022). Certainly, high-level factors can further impact orienting-related PDRs (e.g., Joshi & Gold 2020, Strauch et al. 2022). This is usually reflected in the PDR curve shape, manifesting as a more sustained response (for more, see Strauch et al. 2022). This means that the latency of attention-related PDRs is indicative of the awareness level or processing depth: whereas transient responses occur pre-attentively, more prolonged responses emerge in full awareness (for discussion, see Strauch et al. 2022). In other words, orienting-related PDRs, initiated by an auditory deviant event, can further activate the executive network, making the deviance available to top-down processes or long-term memory. Thus, orienting-related PDRs lead to conscious perception through their transformation into executive-related PDRs. Therefore, the route from pre-attentive to conscious perception of a deviance is also reflected in the time course of pupil responsiveness (an integrated readout of attentional networks in the words of Strauch et al. 2022). One could hence argue that transient, signal-sensitive PDR is an instantiation of involuntary attention switch, while a more prolonged PDR, usually elicited by more complex stimulus properties, is a manifestation of voluntary attention alignment.

These patterns in pupil dilation response align with the findings in neurocognitive research reviewed so far in this chapter, showing that signal-based properties are essential for attracting involuntary attention, while voluntary attention is further activated by more complex top-down processes. In this context, we see a striking resemblance in the elicitation of MMN/P3 responses and PDR: the orienting-related PDR amplitude and latency mirrors MMN activation, while the executive-related PDR latency resembles the P3 elicitation. Many brain studies on the localisation of neural activity in specific brain structures support a direct link between the neurophysiological and pupillary underpinnings of attention orienting. Specifically, evidence from multiple research disciplines proposes the locus coeruleus (LC) system as one of the areas underlying P3 generation (for reviews of studies on the relationship between LC, P3, and the orienting response, see Nieuwenhuis et al. 2005, 2011, Joshi & Gold 2020). In this light, it is worth recalling that the LC system also regulates the PDR elicitation that is related to attention. Further, Alamia et al. (2019), in a recent study investigating the link

between pupillary and neural responses to unattended deviances, found a correlation between PDR and central potentials. Besides pupil size dilation in response to deviant events, they also report deviances evoking the classic MMN neural response. As discussed by the authors, MMN is generated (partially) in the anterior cingulate cortex (ACC), a brain structure also responsible for evoking pupillary responses which is closely connected with LC. It appears thus that in addition to the temporal correlation between the neural and pupillary activity in relation to attentional processes, there is also a spatial relation between the two activities. The regulation of attention-related neural and pupillary responses by the same or connected neural circuits supports the idea that the OR involves a wide range of physiological changes occurring simultaneously across multiple indices. Nevertheless, the localisation of brain activity is beyond the scope of this work, and thus will not be discussed further. The pupillary response, and particularly the PDR, as a psychophysiological underpinning of attention orienting is of particular interest for Chapter 5, where it is revisited and investigated as a proxy of attention orienting in speech.

2.6 Summary

This chapter focused on the composite construct of attention orienting towards auditory deviances by exploring its cognitive and neural underpinnings. First, the origins of the attention orienting concept were reviewed. The orienting response (OR) is defined as a shift in attention initiated by an event deviating in some property from the recent past, finding expressions in various psychological, physiological, neural, motoric, or behavioural indices. In this context, it was further shown that the nature of the deviating property that initiates an OR response elucidates whether the shift in attention is involuntary or voluntary. In this respect, empirical studies concerned with the role of auditory cues in attracting attention were further reviewed, establishing the fundamental role of rising acoustic signals in involuntary attention orienting. In addition, it was shown that involuntary and voluntary orienting emerge at different processing stages (pre-attentive and attentive, respectively), as they are regulated by different mechanisms.

Various different methods used to investigate attention orienting were outlined with a focus on ERPs and pupillometry, two techniques that provide a high resolution of temporal information, and thereby enable the measurement of brain and cognitive activity related to attentional processes over time. Special emphasis was also placed on the MMN/P3 ERP components and the PDR as reflexes

of attention orienting. Specifically, MMN and P3 components were shown to be rigorous correlates of attention orienting, reflecting the route from involuntary to voluntary attention. Similarly, PDR was shown to be a valid index of attention orienting, highlighting a striking resemblance between MMN/P3 elicitation and PDR composites. The link between neural and pupillary activity in relation to attentional processes was illustrated, reinforcing the idea of the OR being expressed in multiple physiological exponents at the same time. In keeping with the focus of this book on attention orienting in spoken language and the relevance of intonation for its function, the next chapter introduces fundamental concepts related to prosody and intonation, and reviews important findings in relation to speech perception and processing.

3 Intonation and its association with prominence

Intonation is a fundamental characteristic of spoken language and is usually described as the *melody* of speech, or tone of voice. Intonation and *prosody* are quite often used interchangeably – not unreasonably – as the two terms are closely interrelated. On the one hand, one could think of prosody as an umbrella term, pertaining broadly to the suprasegmental properties of speech used by speakers and listeners to encode and decode a range of interrelated phenomena (e.g., Arvaniti 2022, Grice et al. 2023), such as:

- i. *linguistic* phenomena, like metrical and rhythmic structure, prosodic phrasing and prosodic prominence,
- ii. *paralinguistic* and *extralinguistic* phenomena, like communicative intentions, attitude, emotions, sociolinguistic identity, gender, sex and age.

On the other hand, intonation is more narrowly related to the melodic part of speech, conveying meaning in a “*linguistically structured way*” (Ladd 2008: 4). This encompasses intonational tones being associated to particular positions in the prosodic structure, organised in discrete phonological categories related to intonational phonology (e.g., Ladd 2008). Although intonational phonology separates the phonological organisation of intonation from its phonetic substance,¹ speech is a continuous signal, meaning that intonation is also expressed in the phonetic record. Fundamental frequency (hereafter, f_0), perceived as pitch, and its modulations have been taken to be the main phonetic exponent of intonation (e.g., Arvaniti 2022). Nonetheless, f_0 modulations interact with other phonetic properties such as intensity (perceived as loudness), segmental durations (perceived as length), as well as spectral and articulatory characteristics (e.g., Ladd 2008). In addition, f_0 also serves other prosodic purposes, such as the encoding

¹It is proposed that this separation enables intonational models to form descriptions and set requirements of intonational analysis at a more abstract level, so that similar levels of analysis can apply across languages (for more, see Ladd 2008, Arvaniti 2022).

3 Intonation and its association with prominence

of lexical contrasts in Mandarin or Cantonese (e.g., Arvaniti 2022), or paralinguistic and extralinguistic information (e.g., Grice et al. 2023). Therefore, it is important to clarify that while f_0 constitutes the central phonetic manifestation of intonation, f_0 and intonation are not synonymous (e.g., Arvaniti 2022).

When we speak, our intonation conveys meaning over and above the meaning of the syllables, words, and sentences bearing it. This means that the intonation of an utterance, also known as the *tune*, can differentiate between a number of meanings, and serve several functions. For example, an intonational tone can indicate that an utterance is incomplete and that there is more to come, rather than being complete, or that it is a question rather than a statement. In addition, speakers make use of intonation to highlight particular information in their utterances or to chunk information into smaller units. Both of these functions, highlighting and chunking, serve to orient listeners' attention to informative parts of utterances which are perceived as *prominent*. As already discussed in the introductory chapter of this book, *prominence* is the property of a linguistic unit "standing out" from its neighbouring environment (e.g., Terken & Hermes 2000, Streefkerk 2002, Himmelmann & Primus 2015, von Heusinger & Schumacher 2019, Cangemi & Baumann 2020).

In the domain of prosody, prominence is referred to as *prosodic prominence* and pertains to entities like syllables, feet, prosodic words, and intonational phrases (e.g., Grice & Kügler 2021). Prosodic prominence is manifested both in discrete phonological choices as well as in continuous physical parameters. Phonologically, prominence is related to elements appearing in "privileged" positions in the prosodic structure (for more on the structural notion of prominence, see Cangemi & Baumann 2020, Grice & Kügler 2021). Phonetically, prominence marking is associated with an enhancement of the physical properties (i.e., signal-based variables) characterising the production of an entity, e.g., f_0 , intensity, duration, periodic energy, and other spectral characteristics (e.g., Grice & Baumann 2007, Baumann & Winter 2018, Grice & Kügler 2021, Albert 2023). In West Germanic languages, a lexically stressed syllable stands out from the rest of the syllables within the same word due to its metrical and/or phonetic characteristics (e.g., Cangemi & Baumann 2020). The prominent syllable at the word level is pivotal in higher prosodic structures like feet or prosodic words, due to the fact that the prominent syllable forms the head of these higher prosodic constituents, which in turn enables a structure to be built around them (e.g., Grice & Kügler 2021). At the utterance level (also called *phrase level*, or *postlexical level*), for instance, a constituent in focus stands out from the remaining constituents in the background due to the type of pitch accent (i.e., postlexical tonal movements associated with the stressed syllable [head]) as well as the enhancement of the phonetic

3.1 Intonational phonology and prosodic structure

exponents of the particular pitch accent it bears (e.g., Bolinger 1986, 1989, Baumann & Schumacher 2012, Lorenzen et al. 2022). Work on prosodic prominence has shown that intonational rises, both in the sense of discrete phonological categories and continuous f_0 shapes, are pivotal cues to prominence perception, and consequently for attracting attention (e.g., Baumann & Röhr 2015, Knight 2008, Niebuhr 2009, Baumann & Winter 2018). The fundamental importance of intonational rises in speech perception and processing is at the core of this book and is further discussed in the remainder of this chapter.

The current chapter offers a concise exploration of concepts related to spoken language, focusing especially on intonation and its association to the notion of prominence. Exploring this association is particularly important for the overarching goal of this book, which is to better understand the relevance of intonation for the orienting response in spoken language, given that the notion of prominence pertains to concepts defining general cognition (including the orienting response; see Chapter 1). The structure of this chapter is the following: Section 3.1 elaborates on the phonological organisation of intonation and its relation to prosodic structure, and discusses one of the most influential accounts of intonational phonology. Section 3.2 discusses the role of intonation in the prosodic marking of prominence and shows that intonational rises are a decisive cue to prominence. Subsequently, Section 3.3 reviews studies supporting the linguistic importance of intonational rises in speech processing. Next, Section 3.4 elucidates the interplay between signal-driven and context-driven processing, in pointing out that signal-based cues can potentially be overwritten by contextual cues, leading to different processes. The chapter closes with Section 3.5, in which list intonation is discussed as the linguistic context of the two experimental studies presented in Chapters 4 and 5.

3.1 Intonational phonology and prosodic structure

Ladd (2008: 4) defines intonation as “the use of *suprasegmental* phonetic features to convey ‘postlexical’ or *sentence-level* pragmatic meaning in a *linguistically structured way*”. The idea that intonation is organised in phonological categories, and is separated from phonetic substance, has been postulated in many theories of intonational analysis, with one of the most influential being the *autosegmental-metrical* framework (henceforth, AM; e.g., Liberman 1975, Pierrehumbert 1980, Ladd 2008).² As discussed in Grice (2022a), the AM framework has been influen-

²AM is not the only framework positing the idea of intonational categoricity. For example, the British School, the IPO model, and other experimental approaches have proposed categorical uses of intonation (for a comprehensive review, see Roessig 2021).

tial for the development of many phonological models of intonation (e.g., Beckman & Pierrehumbert 1986, Shattuck-Hufnagel & Turk 1996, Gussenhoven 2004) as well as psychological models of speech production and perception (e.g., Frazier et al. 2006, Watson et al. 2008, Wagner & Watson 2010, Shattuck-Hufnagel 2020). It has also been central in the advancement of the Tones and Break Indices (ToBI) intonational transcription systems (for a summary, see Beckman et al. 2005) and has formed the basis of prosodic typology (e.g., Jun 2005b, 2014a).

AM theories assume a hierarchical organisation of prosodic structure, wherein the intonational contour, that is the tune of an utterance, consists of distinct units (referred to as *tonal* or *intonational events*) which are independent from the words bearing them, that is, they are postlexical. These distinct units (or else primitives) constitute sequences of H(igh) and L(ow) *tones* which pertain to nodes on the prosodic tree. A prosodic tree involves nodes related to the word level, that is morae, syllables, feet, prosodic words, as well as higher nodes, above the word level, including Accentual Phrases, Intermediate Phrases, and Intonational Phrases. The phonological identity of a tone is mapped onto the continuous acoustic features of its realisation (e.g., Arvaniti 2022). The specifics of the phonological representation and phonetic manifestation of tones are language-specific (e.g., Jun 2005a). Given that the current work focuses on German, let us zoom in on West Germanic languages.

West Germanic languages, such as English, Dutch, and German, are characterised by similar prosody and intonation. For instance, they all have lexical stress and use postlexical tonal events with specific association and functional properties (e.g., Grice et al. 2005). In what follows, I will focus only on postlexical tonal events. According to AM theory and prosodic typology (e.g., Ladd 2008, Jun 2014b), there are two basic ways in which postlexical tonal events can be associated with positions in the prosodic structure in West Germanic languages: as *pitch accents* or as *edge tones*.³ Pitch accents and edge tones have distinct association properties. Whereas pitch accents are associated with stressed syllables, that is, with the head of a constituent, edge tones are associated with initial or final boundaries of smaller or larger constituents. Pitch accents always contain one starred tone, which is the primary tone associated with a tone-bearing unit and marked by an asterisk, or “star” (*). Pitch accents may further contain an additional tone (e.g., H* or L+H*). Edge tones are marked either by the percentage symbol % (e.g., H%) or by the hyphen symbol - (e.g., H-), depending on whether they associate to a larger or smaller unit, respectively. Figure 3.1 illustrates the association properties of an utterance’s tonal events with respect to the nodes

³Hereafter, the terms edge/boundary tone are used interchangeably.

3.1 Intonational phonology and prosodic structure

of its prosodic tree (example and figure are taken from Grice 2022a: 394). From top to bottom, Figure 3.1 illustrates the utterance's Intonational Phrase node (IP), consisting of two intermediate phrases (ip), each of which contains one or two accentual phrases (φ). The smallest level of phrasing here consists of one or more prosodic words (ω). In turn, a prosodic word consists of one or more feet (Σ), and a foot consists of one or more syllables (σ). In the bottom of Figure 3.1, we can see the tonal structure associated with the utterance. Specifically, there are four starred tones (see orange boxes), i.e., pitch accents, associated with the stressed monosyllabic words *too*, *cooks*, *spoil*, and *broth*. These four starred tones dock onto the head of a foot, respectively. Three further tonal events, marked in green boxes, pertain to the two ip and the highest IP nodes, respectively, forming the edge tones which mark both smaller and larger units in the structure.

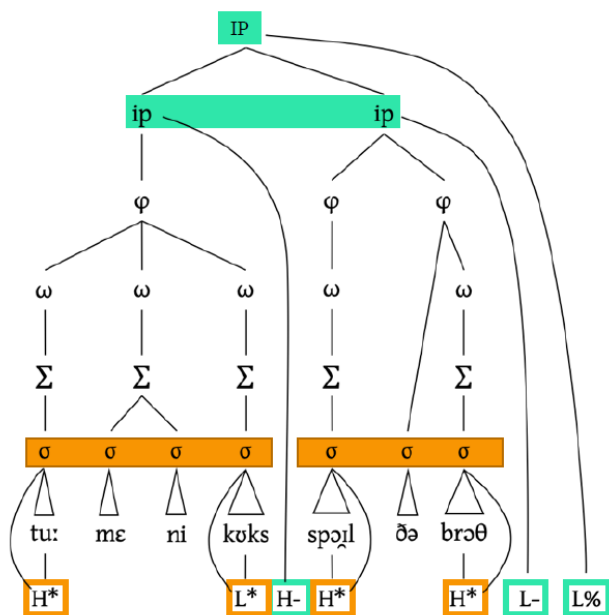


Figure 3.1: Prosodic hierarchy of the utterance “Too many cooks spoil the broth”. Figure reproduced from Grice (2022a: 394).

In AM, functional properties also play a role in the distinction between pitch accents and edge tones. According to AM and prosodic typology (e.g., Ladd 2008, Jun 2014b), pitch accents not only associate with stressed syllables, but they are also considered to primarily have a highlighting function related to prominence-cueing in West Germanic languages, which make use of both pitch accents and

3 Intonation and its association with prominence

edge tones.⁴ By contrast, edge tones not only associate with initial or final edges of smaller or larger constituents, but it is claimed that they are mainly used for phrasing and juncture-cueing. Thus, it appears that the association properties of tones (accent/boundary) are pre-packaged with distinct functions, as reported in Grice (2022a): pitch accents cue prominence; boundary tones cue phrasing. Nonetheless, a growing number of research on both speech production and perception challenge this pre-defined relation between association and functional properties of tones postulated in AM models. For example, as discussed in Grice (2022a,b) tones in Tashlhiyt Berber have no strict association properties, neither to heads nor to edges of constituents. Nevertheless, these tones appear to have a prominence-cueing function, since they are associated with words in focus. The loose association properties of tones in Tashlhiyt Berber implies that they are not pitch accents, given that they do not align with a prosodic head, although they do serve the highlighting function attributed to pitch accents (for more on Tashlhiyt Berber, see Bruggeman et al. 2017, Bruggeman 2020). Further, Grice (2022a,b) discusses the case of Maltese: whilst Maltese is described as an intonational language having regular pitch accents (in terms of AM), it commonly exhibits additional initial peaks on unstressed syllables. There is evidence that these initial peaks do not serve a demarcating function, although they are associated with the left edge of the word. It mostly appears that words in Maltese can have two types of prominence, that is, lexical stress prominence, and pitch prominence on the initial syllable. It seems to be the case, then, that tones in Maltese can associate with both edges and heads of words, but that their association properties do not necessarily lead to the assumed AM functions of demarcating or prominence-cueing (for more on Maltese initial peaks, see Vella 1995, 2011, Grice et al. 2019, Lialiou et al. 2021, 2023).

This evidence calls into question the link between the association (head/edge) and functional properties (prominence/phrasing) of tonal events postulated by AM, and highlights the need for a dissociation of the two. Such a dissociation will allow for more flexibility in describing diverse phenomena. This is necessary not only for “new” or under-studied languages but also for (re)describing phenomena in already well-studied languages. Section 3.3, reviewing work on well-studied head prominence languages, shows that edge tones can, in fact,

⁴One of the three criteria for language classification in Jun’s typology is the *prominence type*. Languages which cue prominence by marking the head of a constituent are known as *head prominence* languages. Languages which cue prominence by marking both the head and the edge of a prosodic constituent are known as *head/edge prominence* languages. Finally, languages that lack both lexical and postlexical heads and cue prominence by marking the edge of a constituent are known as *edge prominence* languages.

also affect prominence perception. Prior to that, Section 3.2 discusses the link between intonational rises and prominence.

3.2 Intonational rises as a prominence-cueing device

As already mentioned at the beginning of this chapter, intonation serves multiple functions, from helping to locate fixed attributes of words – such as lexical stress in West Germanic languages – to prominence marking at the utterance level (for a comprehensive review, see Grice et al. 2023). Speakers may use a wide assortment of linguistic means for signalling prominence, including prosodic and non-prosodic factors (e.g., Baumann & Winter 2018). Nonetheless intonation is a pivotal cue to prominence in spoken communication. Specifically, speakers make use of intonation to highlight important parts of their message, endeavouring to orient listeners' attention to specific information. At the same time, intonation is crucial during speech processing, since prominent intonational events attract and allocate listeners' attention to the constituents bearing them, facilitating their processing. The encoding and decoding of prosodic prominence relates to discrete phonological choices, continuous phonetic modulations reflected in the speech signal, and to meaning derived from the context (see Chapter 1, and Section 3.4 in the current chapter).

Whereas the specifics of prominence marking using intonation depend on a language's prosodic system (e.g., Rosenberg et al. 2012, Smith & Rathcke 2020), cross-linguistic research has revealed f_0 to be one of the primary cues (e.g., Liberman 1975, Pierrehumbert 1980, Ladd 2008, Beckman 2012). Specifically, in West Germanic languages, modulations in f_0 direction, excursion, scaling, and timing are some of the acoustic dimensions found to be indicative of different degrees of perceived prominence, such that f_0 peaks are generally perceived as more prominent than f_0 valleys (e.g., Rietveld & Gussenhoven 1985, Gussenhoven et al. 1997, Ladd & Morton 1997). What is more, the size of the excursion has been found to be a decisive cue to prominence: the steeper the f_0 movement, the more prominent the word associated with it (e.g., Hart et al. 1990, Gussenhoven & Rietveld 1988, Gussenhoven 2004). Further, it has been shown that the general shape of the f_0 contour (e.g., the steepness of a rise or fall as well as its close [early/late] alignment with a stressed syllable) is pivotal to prominence (e.g., Kohler & Gartenberg 1991, Niebuhr 2009, Knight 2008).

3 Intonation and its association with prominence

Before discussing the phonological expressions of prominence marking, it is important to note that although f_0 modulations are central to prominence marking postlexically, modulations of other phonetic exponents are also relevant. Perceived prominence has also been found to correlate with enhanced intensity, increased duration, enhanced spectral characteristics, articulatory effort, and, in more recent work, enhanced periodic energy (e.g., Beckman & Edwards 1994, Cole et al. 2010, Baumann & Winter 2018, Roessig & Mücke 2019, Albert 2023). It appears therefore that a multiplicity of cues interact with each other in complex constellations to express postlexical prominence. Most importantly, research on individual variability has shown that speakers encode and decode prominence relations using the aforementioned features in different combinations and with different degrees of strength (e.g., Baumann & Winter 2018, Lorenzen et al. 2023, 2024).

With respect to the phonological choices of postlexical prominence marking, in languages like German (a West Germanic *head* prominence language according to Jun's typology), prominence is encoded (and decoded) through the use of pitch accents. That is, a highlighting function is attributed to tonal movements associated with the stressed syllable (head; for more, see Section 3.1). As mentioned in the previous section, the identity of a pitch accent (and any other postlexical tone) is defined by its phonetic substance. Specifically, the realisation of pitch accents in the acoustic record results in different pitch accent types. Accent types are defined on the basis of the aforementioned f_0 dimensions; f_0 direction (rise/fall), f_0 scaling (steep/shallow), f_0 height (peak/valley), and f_0 alignment (timing of f_0 peak or valley relative to a stressed syllable). Research on prominence perception in West Germanic languages has shown that the type of pitch accent impacts the degree of perceived prominence.⁵ For instance, studies on Dutch (e.g., Rietveld & Gussenhoven 1985, Gussenhoven & Rietveld 1988), English (e.g., Ladd & Morton 1997, Knight 2008, Cole et al. 2019), and German (e.g., Kohler & Gartenberg 1991, Niebuhr 2009, Baumann & Röhr 2015, Baumann & Winter 2018) have found that the overall shape of an accent/ f_0 contour affects prominence perception, revealing the importance of intonational rises, as higher f_0 peaks and steeper f_0 rises are perceived as more prominent.

Let us now zoom in further on intonational rises and their prominence status

⁵Pitch accent placement within an utterance (prenuclear, nuclear, postnuclear) is also an important factor in prominence perception. Many schools of intonational analysis have claimed that the last accent in an utterance, the *nuclear accent*, is the most prominent one. For German, the following prominence hierarchy of pitch accent placement has been proposed: nuclear > prenuclear > postnuclear (e.g., Baumann & Röhr 2015, Grice et al. 2017). Pitch accent placement is beyond the scope of this book, and will therefore not be discussed further.

3.2 Intonational rises as a prominence-cueing device

in German. Baumann & Röhr (2015), in a prominence rating task, investigated the degree of accent type prominence using the accent inventory of the German Tones and Break Indices model (GToBI; see Table 3.1; Grice et al. 2005). The items of the study consisted of utterances featuring proper names. The authors manipulated the pitch accent contour on the proper names, and instructed the listeners to rate how prominent a given proper name sounded using a visual analogue scale. The results of this study show that the accent type on a given proper name affected its perceived prominence: when proper names were produced with pitch rises, they were perceived as more prominent than when they were produced with pitch falls (f_0 direction). Crucially, the rise or fall occurred on the stressed syllable of the proper name. In a similar fashion, steeper rises and falls on proper names were rated as more prominent than shallow rises and falls (f_0 excursion), and finally, pitch peaks on proper names were judged as more prominent than pitch valleys (f_0 height). Table 3.1 presents the perceived prominence hierarchy of the tested German accent types proposed by Baumann & Röhr (2015).

Table 3.1: GToBI accent type prominence hierarchy based on the three dimensions: pitch direction, excursion, and value of starred tone. Increasing prominence is ranked from bottom to top. Table reproduced and slightly adapted from Baumann & Röhr (2015: 2).

Accent type	Pitch direction	Pitch excursion	Value of starred tone
L+H*	rise	steep	H
L*+H	rise	steep	L
H*	rise	shallow	H
!H*	rise	shallow	!H
H+!H*	fall	(relatively) steep	!H
H+L*	fall	steep	L
L*	level	shallow	L
∅	n/a	n/a	n/a

In a more recent study, Baumann & Winter (2018), using a Rapid Prosody Transcription (RPT) task, tested the impact of continuous prosodic (e.g., f_0 , duration, intensity), discrete prosodic (pitch accent type and placement), as well as non-prosodic (e.g., lexical, syntactic, semantic) variables on prominence perception. Specifically, in this RPT task, untrained German listeners were asked to underline words regarded as prominent during or right after listening to an utterance. The results show that discrete prosodic variables, including pitch accent type, were

3 Intonation and its association with prominence

the best predictors of prominence. More specifically, the particular shape characteristics of the pitch contour, with rising contours being pivotal, were found to be the primary cue to prominence perception. This study also highlighted a new – at the time – perspective related to prominence perception. In particular, the authors mention that despite the overall trend found at the group level, prominence perception in their sample was partly determined by individual differences, with some of the listeners attending mostly to pitch-driven cues, and others exhibiting a reverse pattern in attending mostly to lexical and semantic cues.

So far, it has been shown for West Germanic languages, including German, that discrete (phonological) and continuous (phonetic) prosodic cues pattern together in flagging postlexical prominence, reflecting the fact that the shape of a pitch contour is defined by its f_0 dimensions. Summarising,

- i. rises are more prominent than falls,
- ii. steep rises are more prominent than shallow rises and falls, and
- iii. peaks are more prominent than flat pitch and/or valleys.

The results discussed in this section strongly suggest that intonational rises constitute a decisive cue in prominence perception. The next section elaborates further on the linguistic importance of intonational rises in speech processing.

3.3 Intonational rises in speech processing

It was pointed out in Section 2.4 that in auditory processing, the physical properties of the acoustic signal are essential for attracting attention. For instance, rises in the amplitude or pitch of sine waves prompt auditory looming effects, indicating that the sound source is approaching. Acoustic saliency thus serves as a warning cue that activates attentional resources (e.g., Bach et al. 2008). Falling acoustic signals may also attract attention, but it has been shown that these are experienced mostly as fading (indicating a receding sound source), and are thus processed differently than rises (e.g., Macdonald & Campbell 2011).

In a similar vein, related research has shown that prosodic cues, such as intonational rises, are pivotal in speech perception and processing. In West Germanic languages, rising pitch accents typically mark important and informative constituents in an utterance, such as new topics, new referents, or different types of focus (especially contrastive or corrective focus; e.g., Bolinger 1986, 1989, Baumann & Schumacher 2012, Grice et al. 2017, Lorenzen et al. 2022, 2024). Rising

3.3 *Intonational rises in speech processing*

intonation allocates attention towards the constituents bearing it, which facilitates their processing. Further, rising pitch accents, when used as a focus-cueing device, also guide attention towards semantic incongruencies, because they lead to more elaborate processing of the focused information (e.g., Ventura et al. 2020, Wang et al. 2011). Moreover, intonational rises guide attention in memory and recall tasks, facilitating recall performance (e.g., Fraundorf et al. 2010, Koch & Spalek 2021, Röhr, Savino, et al. 2022). Last but not least, information highlighted by rising intonation is perceived as more prominent, which leads to more attentional resources being allocated to information highlighted in this way than when it is highlighted by falling intonation (e.g., Sullivan & Horowitz 1983, Hsu et al. 2015, Röhr et al. 2020, 2021). Nonetheless, not all rising pitch accents attract attention. For instance, Bari Italian postfocal rising pitch accents, which are used to mark interrogatives, do not attract attention (i.e., postfocal rising pitch accents do not facilitate detecting incongruencies in the postfocal domain; Ventura et al. 2020, Ventura 2020). It thus appears essential to take the structure of the language under investigation into account. In West Germanic languages, prominence is very often encoded in the form of intonational rises. Therefore, the use of rising accents to mark important constituents in such languages may be the reason for their attention orienting potential.

Let us now focus on the triplet intonation-attention-speech processing. It appears that the auditory and linguistic processing systems share many common features, especially with regard to prediction and attention orienting. These two mechanisms appear to be driving forces for information processing in general, including for linguistic input (see Bornkessel-Schlesewsky & Schumacher 2016). As laid out in Section 2.3, the auditory system develops sensory-memory representations related to the auditory environment, on the basis of which it predicts the properties of the upcoming sound events. When predictions are not attested, attention-calling mechanisms are activated. Likewise, according to Schumacher et al. (2015) the linguistic processing system constantly develops mental representations from which predictions for the upcoming entities are derived (for more detail on the general architecture of the prediction-updating cycle, see Röhr et al. 2021, Repp & Schumacher 2023). As already mentioned in Chapter 2, according to Friston (2010) our brains build an internal model of the world on which we rely to predict upcoming sensory events. When these predictions are violated, prediction errors – to use Friston’s term – arise. Prediction errors in language, namely “mismatches between context and target items” (Röhr et al. 2021: 36), have been found to elicit detection and attention-calling processes as well. In the previous chapter, we saw that after an auditory deviation/mismatch, the first attention-calling mechanism that is activated in the auditory cortex is the mismatch nega-

3 *Intonation and its association with prominence*

tivity (MMN) response, oftentimes called the regularity-violation response (e.g., Näätänen 1992). Mismatches in language function in a similar way, eliciting negative neurophysiological deflections related to such violations (for a general account on negativities as reflexes of prediction errors, see Bornkessel-Schlesewsky & Schlewsky 2019). The most well-studied prediction error deflection in language is the N400. The N400 “represents processes that affect an early discourse linking stage and derive from expectations for an entity raised by the discourse context. The higher the processing cost for an entity or the more difficult its reconciliation with prior discourse, the more enhanced is the N400 amplitude” (Röhr et al. 2021: 36). Röhr et al. (2021) point towards ERP studies on pitch/intonation processing, reporting that as pitch unfolds in time, listeners incrementally develop expectations for the upcoming entities and their prosodic realisation. These studies report that when a mismatch between the prosodic input and the upcoming information occurs, an N400, indicative of a prediction-violation (for more, see Chapter 2), is elicited in the brain (e.g., Toepel et al. 2007, Baumann & Schumacher 2012). As I reviewed in the previous chapter, attention orienting towards an auditory mismatch in cognition can also elicit a positive deflection – identified as part of the P3 family – which involves a conscious perception of this mismatch. Likewise, Röhr et al. (2021) investigated the signal-driven effects of prosodic prominence in an EEG study and found that intonational rises, by virtue of being acoustically prominent, attracted more attention than less prominent intonational patterns, eliciting a P3 response (for a detailed review of this study, see Section 3.4). Further, Röhr et al. (2021) mention that attention orienting in language can also be reflected in a Late Positivity, also identified within the P3 family (e.g., Ruchkin et al. 1990: see also Chapter 2). Röhr et al. (2021) propose that Late Positives in language reflect an update of the mental model. More specifically, according to Röhr et al. (2021), listeners, after orienting their attention towards a prosodic mismatch, topic shift, or novel information, need to incorporate the new knowledge in their mental representations by updating their mental model. Late Positivity has also been considered as an index of attention orienting in studies investigating prosodic cues marking new information and focused constituents (e.g., Toepel et al. 2007), as well as unexpected accentuation (e.g., Schumacher & Baumann 2010, Baumann & Schumacher 2012), redundant accentuation on constituents in the background (e.g., Li et al. 2018), and deviant or unexpected prosodic patterns (e.g., Kakouros et al. 2018). According to Röhr et al. (2021), these are cases where a listener perceives a conflict between prosodic and information structural cues. This conflict can be resolved through the updating processes, whereby the mental model is repaired and reorganised (see also, Toepel et al. 2007, Dimitrova et al. 2012, Brouwer & Hoeks 2013, Schumacher et al. 2015).

3.3 *Intonational rises in speech processing*

Evidence supporting the privileged position of pitch rises in attracting attention also comes from studies on the neural processing of linguistically meaningful pitch variations at both the lexical and postlexical level in Mandarin Chinese (e.g., Ren et al. 2009, Tsang et al. 2011, Liu et al. 2016, Li & Chen 2018). These studies show that brain responses are not only activated by the acoustic contrasts in the signal, but they are also sensitive to the timing of the acoustic cues. The study by Li & Chen (2018) is of particular relevance for the current work, as it investigated Tone2 (that is, a rising tone) and Tone4 (that is, a falling tone), similarly to Chapter's 4 intonational rising and falling contours. Using the MMN paradigm (for more on this paradigm, see Chapter 2), the authors compared the timing of the cues by contrasting Tone2/Tone4 with Tone3 as a reference tone, and they found that MMN is sensitive to the timing of the acoustic cues (i.e., the time point when the tone contours start to diverge; cue of divergence point). It is particularly interesting that the MMN time window they found for the contours with the early divergence cue is the same time window observed in the current study reported in Chapter 4. The contours used in the study reported in Chapter 4 also have an early divergence point (i.e., divergence from the beginning of the word), pointing at a cross-linguistic similarity.

As discussed in Section 3.1, AM theory and prosodic typology postulate that tonal events come pre-packaged with specific associations and functional properties (in the case of English, German, and other languages with a similar prosodic structure). More specifically, pitch accents are assumed to cue prominence, while edge tones serve to chunk utterances into smaller units (e.g., Ladd 2008, Jun 2014b). It is therefore proposed that accentual rises constitute a better cue in directing listener attention than rises at prosodic boundaries. Nevertheless, studies investigating prominence perception and processing call into question the strict dichotomy proposed by the AM theory. Specifically, it has been claimed that prosodic phrasing actually can affect prominence perception (for discussion, see Grice & Kügler 2021). Some evidence that rising edge tones can also cue prominence comes from serial recall tasks of nine-digit sequences in Italian (Savino et al. 2020) and German (Röhr, Savino, et al. 2022, Grice et al. 2024). These studies found that boundary rises marking the last item of non-final triplets facilitated the recall accuracy of the digit in the boundary position that bears the edge tone. This facilitation was also observed over the whole triplet. These results reveal that edge tone rises appear to cue prominence on the whole domain that is delimited. Other evidence for edge tones cueing prominence comes from processing studies on the domain of the word (for discussion, see Grice & Kügler 2021), showing that rising boundary tones facilitate, for example, word segmentation (e.g., Ou & Guo 2021) and word recognition (e.g., Kember et al. 2021).

3 Intonation and its association with prominence

Therefore, regardless of the functions with which pitch accents and boundary tones are pre-defined within AM theory, the prominence patterns in a language can be affected or modulated by different structural positions (for the notion of structural prominence, see Streefkerk 2002, Himmelmann & Primus 2015, Cangemi & Baumann 2020). Put differently, AM postulates that prominence in West Germanic languages is mostly expressed in the form of pitch accents, directing the listener's attention to the highlighted information. The stressed syllable, as the docking site for the pitch accent, is the head of the word, and therefore occupies an essential position in the linguistic structure. Yet, prosodic boundaries appear to also guide attention towards important elements. This could be so because flagging prominent information at privileged positions, such as in the beginning or at the end of an utterance, is crucial for speech processing and planning (e.g., Seidl & Johnson 2006, Ou & Guo 2021). Hence, one could argue that in a complex signal such as speech, prominence might not necessarily be encoded by cues on one specific structural position, but rather by the combination of cues on different privileged positions. Both experimental studies reported in this book (Chapters 4 and 5) investigate the role of the phonological status of the rise, namely as pitch accent or edge tone, in attention orienting. The key, and novel, discovery is that rises associated with the edges of constituents can also direct the listener's attention towards the constituents bearing them; the evidence comes from ERP as well as pupillometric data.

Although linguistic research has attested that rising pitch is pivotal in spoken communication, previous research on the processing of unexpected/deviant auditory events has focused on pure tones only. To my knowledge, the only oddball study that has explored the sensitivity of pitch change direction in relation to speech is Hsu et al. (2015). However, although some of Hsu et al.'s stimuli were produced with a human voice, they were not designed to convey any linguistic meaning (see the review of this study in Section 2.4). Chapters 4 and 5 delve into the role of intonational rises in attention orienting, using speech stimuli that convey linguistic meaning. Specifically, in extending previous neurophysiological work on simple sine waves, the study reported in Chapter 4 introduces rising and falling intonation (attributable to both accents and boundary tones) on sequentially presented lexical items, yielding a list context and the concordant, language-specific expectations. Likewise, Chapter 5 uses domain-final rises and falls (i.e., edge tones) within a numerical list context. Crucially, the rise or fall for the stimuli used in the Hsu et al. paper involved a change in pitch from one stimulus to the next (i.e., across stimuli), whereas in the current work the rise or fall takes place within the stimulus itself.

3.4 Signal- and context-driven processing

The processing of information is highly affected by listener expectations (e.g., Clark 2013, Huettig 2015, Grice et al. 2017, Friston 2018, Roessig et al. 2019). These expectations are in part driven by pure acoustic properties, i.e., the prosodic context, but can also be driven by context- and/or language-specific expectations, i.e., expectations raised by the pragmatic context or by language structure. For example, a particular accent on a constituent or an inserted pause can create expectations as to the upcoming information.

Previous research on acoustic mismatch detection has targeted signal-based expectations, but context may also shape attentional orientation. For instance, Röhr et al. (2021) investigated the role of signal- and expectation-driven effects of prosodic prominence in two EEG studies. First, four different German accentual contours were tested in isolated sentences (steep rise, shallow rise, [steep] fall, and no accent) with intonation the only factor influencing attention orienting. Second, the most prominent steep rising accent and the less prominent falling accent were tested with regard to expectations of how exciting/unusual the content of an utterance is, by relating the stimuli to an exciting/unusual and neutral (negligible/ordinary) pre-context. Results indicate that attentional cues, both signal-driven and context-driven, engender positivities of varying latency: (i) a prominent rise on the stressed syllable (and not a rise elsewhere) consumes attentional resources at an early processing stage (reflected in an Early Positivity), and (ii) highlighting induced by the context (i.e., the exciting context) consumes attentional resources at a later processing stage (reflected in a Late Positivity). Moreover, results show that prior context builds up expectations for the upcoming prosodic input, reflected in N400 prediction errors engendered by acoustically unexpected (here, prominent) accents as well as contextually inappropriate prosodic realisations. These studies hence suggest that attentional orientation and predictive processing reflect discrete stages in the construction of a mental representation during real-time comprehension. Another example of the interaction between signal- and context-driven processing of pitch variation is the study by Liu et al. (2016). Liu and colleagues investigated the attentive processing of tone and intonation in Mandarin Chinese and found that conscious attentional neural responses (i.e., the P300 response) were modulated by the pragmatic context (question vs. statement) in which Tone4 appeared.

Following up on Röhr et al. (2020, 2021), a recent study by Röhr, Baumann & Grice (2022) on German investigated the effect of context-driven expectations on tonal cues to prominence. Specifically, the authors tested contexts where the speaker needs to mark information in an exciting way (exciting context) as op-

3 Intonation and its association with prominence

posed to contexts where the speaker provides information in an ordinary way (neutral context). For the former context, the authors found that speakers made the target information prominent, by using mostly rising accents characterised by large rising tonal onglides (measuring the pitch excursion between the pre-stressed syllable and the accentual peak). In contrast, speakers marked ordinary information with less prominent accents, mostly by using small rising or falling onglides. Listeners interpreted large rising onglides (as in L+H*) as unusual/exciting/prominent information, as opposed to small rising onglides (as in H*) and falling onglides (as in H+L*). This study provides further evidence not only for the intrinsic association between rising intonation and prominence, but also for the interplay between signal-driven and context-driven cues.

In addition, studies have shown that the processing and perception of prosodic prominence can also be shaped by language-specific expectations, such that attention is allocated to different information based on a language's prosodic structure (e.g., Chandrasekaran et al. 2009, Ventura et al. 2020). For example, the degree of prosodic attenuation in postfocal material in Italian is different than in German. In parallel collaborative work with Grice et al. (2025), we conducted a web-based experiment investigating how native Italian speakers and German learners of Italian perceive the prominence of words with different focus structures (i.e., in broad and narrow focus, and postfocally) in Italian. In this study, we were particularly interested in the question of how listeners' native language shapes the perception of postfocal prominence. The results show that both groups of listeners perceived words in narrow focus as more prominent than words in broad focus, but that they differed in how they perceived postfocal words. Specifically, native Italian listeners perceived words in both broad focus and postfocal position equally prominent, despite the differences in the acoustic realisation of the target words in the two focus conditions. In contrast, the German learners perceived postfocal words as less prominent than words in broad focus. These findings indicate that native Italian speakers used both signal- and language-driven expectations in their perception of postfocal prominence, whereas the prominence perception of German learners appeared to be driven only by the signal, i.e., the acoustic saliency of the broad focus words compared to the reduced acoustic properties of the postfocal words. Both Grice et al. (2025) and Ventura et al. (2020) show that flat pitch appears to have a different status in Italian and German, as in Italian attention to postfocal material is enhanced by flat pitch. Therefore, German learners in Grice et al. (2025) appeared to also use their L1 expectations regarding the deaccentuation of postfocal elements.

In sum, it appears that prior expectations driven by the linguistic or discourse context, or even by recent speech experience, can overwrite the typical acous-

tic cues to prominence (e.g., Cole et al. 2010, Bishop 2013, Kakouros et al. 2018, Roettger & Rimland 2020), leading to different processes: whilst signal-driven cues are decisive mainly for pre-attentive processes, context-driven cues mainly influence attentive processes. This point is revisited and discussed in depth in Chapter 4.

3.5 A brief note on German list intonation

The previous sections of this chapter established the relevance of intonation for prominence. It is worth keeping in mind, however, that intonation serves multiple other functions as well. Crucially for the present work, intonation can differentiate between continuity and finality in utterances and – most importantly for this work – in lists. Thus, before closing this chapter, let us touch upon German list intonation, the linguistic context of the two experimental studies reported in this work.

German list intonation involves different pitch contours depending on whether the list items are non-final or final. More specifically, rising and level⁶ contours are typically used by native speakers to recite elements of a “multi-part unit such as a list...” (Peters 2018: 98), as rising and/or level contours on non-final and penultimate items in lists (or in utterances in general) denote continuity. In the context of a list, these contours indicate that the list is open or incomplete (e.g., Grabe 1998, Baumann & Trouvain 2001, Chen 2003, Peters 2018). The final element in a (closed) list, by contrast, is typically marked by a falling contour, indicating finality and marking the end of the list (see Baumann & Trouvain 2001, Peters 2018). Nonetheless, final rises (the reflex of H% boundary tones) can also mark the end of units, such as a sublist, as there are different *types* of continuation. Continuation can be signalled both across smaller or larger domains (e.g., Chen 2003). In discourse, this translates to sentence-internal continuation and continuation at clause-boundaries (e.g., Chen 2003), while in the case of a list, this corresponds to list-internal (i.e., non-final items) and list-boundary (i.e., end of a sublist where a subsequent list or another utterance is yet to come) continuation.

Figure 3.2 presents stylised examples of some possible intonational contours marking non-final and final items in a numeric list. In blue, we see instantiations of different rising (first and second row), high (third row), and low level (final row) contours that can mark non-final items in a list, indicating list-internal continuity.

⁶With level contours, I refer to contours that are flat or shallow falling towards the boundary. The pre-boundary tonal specification can vary from rise-fall to shallow falling or shallow high.

3 Intonation and its association with prominence

In red, we see contours that can mark final items in a list. More specifically, whilst in the first, second, and last row, we see instantiations of different falling contours denoting the end of a closed list, in the the third row, we see, in pink, an example of a final rise (i.e., a boundary rising contour), marking the end of a list and indicating list-boundary continuation.

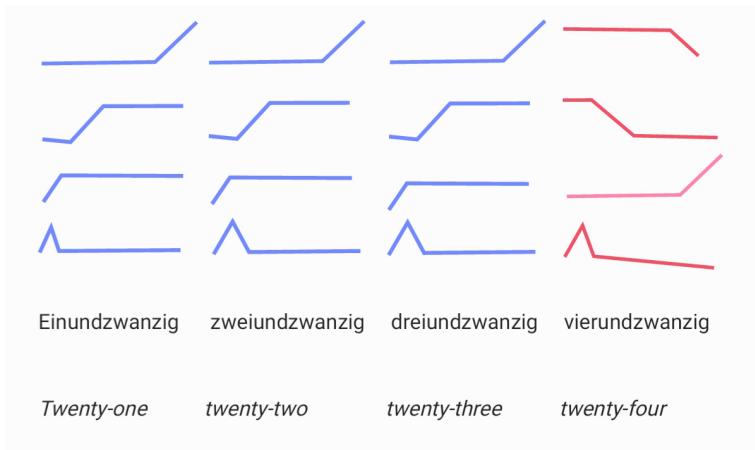


Figure 3.2: Illustrative examples of list intonation in German on non-final and final elements in lists using stylised contours (adapted from examples (14), (15), and (16) in Peters 2018: 98–99).

3.6 Summary

The current chapter served as a condensed introduction to prosody, intonation and prosodic structure. Central for this chapter has been the relevance of intonation for prominence. The studies reviewed and discussed show that rising intonation not only constitutes a decisive cue to prominence, but that it also plays a crucial role in speech processing. However, we saw that, in addition to signal-driven cues (like intonational rises) context-driven (i.e., meaning- or language-specific) cues are also of great importance in speech processing, as they can overwrite the importance of the signal at times, which in turn can lead to the activation of different processes. Here we note an interesting parallel between speech and auditory processing. As we saw in Chapter 2, saliency, both signal-driven and meaning-based, is the mainspring for attention orienting in auditory cognition. In a similar vein, this chapter showed that prominence, both signal-driven and context-driven is a driving force for speech processing. As the

introductory chapter put forward, prominence pertains to the concept of attention (introduced in Chapter 2). We saw in this chapter that the linguistic domain uses a pool of mechanisms shared with general cognition, mechanisms which are crucial for the orienting response in both auditory and speech processing. I will revisit this argument in the upcoming chapters. More specifically, the experimental work presented in Chapters 4 and 5 builds on both neurocognitive and linguistic research in an endeavour to unravel the relevance of intonation as a prominence-cueing device for attention orienting in spoken language.

4 Unravelling the role of intonation in attention orienting using ERPs

One of our fundamental abilities is that, based on our interests and goals, we can voluntarily select our object of attention among the many available input sources. Voluntary attention thus shields the processing of information that we have chosen to process from other available but currently irrelevant information. However, unexpected sound changes outside of the current attentional focus may break through this shield, causing an involuntary switch of attentional resources towards these unexpected changes (see Chapter 2). One of the auditory cortex change detection mechanisms that triggers involuntary attention is reflected in the neurophysiological response called mismatch negativity (MMN; e.g., Näätänen 1992). The present study is concerned with the processing of intonational rises and falls in German when presented unexpectedly in a stream of repetitive auditory stimuli. In an EEG study, using an auditory passive odd-ball paradigm, I investigate whether rising pitch in speech is special in attention orienting by measuring MMN responses to intonational changes.¹

4.1 Motivation

Chapter 2 provided a comprehensive review of studies attesting that the physical properties of the acoustic signal are crucial for the activation of the orienting response, showcasing the orienting function of rising signals (see Section 2.4). Specifically, rises in the amplitude or pitch of sine waves form intrinsic warning cues with ripple effects, prompting auditory looming (i.e., the percept of an approaching sound source), which activates attentional resources and in turn leads to a series of neural, psychological, or physiological reflexive responses. However, it has also been shown that falling acoustic signals, too, may attract attention when presented unexpectedly, even though they are experienced mostly as a receding sound source and thus processed differently than rises. In a similar fashion, Chapter 3 highlighted the linguistic importance of intonational rises in

¹The research presented in this chapter has been reported in Lialiou, Grice, et al. (2024).

4 *Unravelling the role of intonation in attention orienting using ERPs*

speech as a decisive cue to prominence (see Section 3.2), indicating their crucial role in speech processing (see Section 3.3). Speakers make use of prominent intonational contours to highlight important elements in their utterances, endeavouring to orient listener attention to specific information or a specific idea. At the same time, prominent intonational events like rising contours attract and allocate listener attention towards constituents bearing these intonational events, facilitating their processing. However, Chapter 3 showed that, in addition to signal-driven cues (such as rises in pitch), context-driven cues take on a crucial role in speech processing, and sometimes overwrite the typical acoustic cues that signal prominence.

In the current chapter, I revisit the idea of an attentional bias (essentially serving as a warning cue) towards sounds with rising as opposed to falling acoustic properties and extend this idea from general cognition to spoken language. In the context of previous neurocognitive studies suggesting that unexpected sound events with rising acoustic properties attract more attention than sound events with falling acoustic properties in a stream of repetitive non-linguistic auditory events (see Section 2.4), the current study explores the role of intonational rises and falls (attributable to accents and boundary tones) in attention orienting, using speech stimuli that convey linguistic meaning. Extending previous work on simple sine waves, I introduce rising and falling intonation on sequentially presented lexical items, which potentially gives rise to a list context and the concordant, language-specific expectations. In addition, I investigate the role of the phonological status of the rise (pitch accent or edge tone)² in attention orienting. Crucially, intonational theories such as the autosegmental-metrical theory (AM; e.g., Ladd 2008) postulate a strict link between the association (head/edge) and the functional (prominence/phrasing) properties of tonal events. Pitch accents are thereby related to prominence and edge tones to phonological phrasing, excluding a prominence-cueing function of the latter (see Section 3.2 for more detail). However, recent research has provided some evidence against such strict postulations (for review, see Chapter 3). Therefore, the question I address in relation to the phonological status of rises is whether only rises on pitch accents play a role, or whether rising edge tones have a similar effect, as in the serial recall results reviewed in Section 3.3.

For this study, EEG recordings of listeners were collected and event-related brain potentials (ERPs; for more on this method, see Section 2.5) related to unexpected auditory deviances in a repetitive auditory stream were measured. Of

²As already stated in the introductory chapter, the term edge tone is used interchangeably with the term boundary tone.

particular interest for this study are the MMN and the subsequent positivity (P3). Previous studies have shown that these two brain responses appear to be the neurophysiological underpinnings of involuntary and voluntary attention, indexing the path from pre-attentive to conscious processes related to an unexpected auditory change (see Section 2.5). The MMN can be recorded in different stimulus paradigms, depending on the different aspects of the central auditory processing under study. One of the most frequently used paradigms is the classic oddball paradigm in passive recordings, utilised in the current study (e.g., Näätänen et al. 2019). In this paradigm, participants are presented auditorily with sequences of repetitive sounds, called standards, occasionally interspersed with a rare sound, called deviant, during a film with no sound. This allows for the elicitation of brain responses in the absence of attentive listening, and can thus shed light on the nature of the underlying neural mechanisms.

The main question this study puts forward is whether intonational rises attract more attention than falls (RQ1 [rises]). Two central hypotheses are tested. Following the auditory looming literature (e.g., Bach et al. 2008, Macdonald & Campbell 2011), Hypothesis 1 states that unexpected rises in a sequence of repetitive falls should attract more attention than unexpected falls in a stream of repetitive rises, evoking a more pronounced MMN potentially followed by a P3 response. Furthermore, given the focus of the current work on the processing of intonational rises and falls in both accentual and boundary positions, I subsequently ask whether the phonological status of the rise and fall affects the underlying processing (RQ2 [edges]), or in other words whether accentual contours attract more attention than boundary contours. Hypothesis 2 thus states that the extent of the effect may vary as a function of the phonological association (pitch accent vs. boundary tone) of the tonal event (e.g., Ladd 2008). If the standard AM account holds, postulating pitch accents as the primary markers of prominence, a greater MMN effect for accentual over boundary contours is expected.

4.2 **Methods**

Employing the classic oddball paradigm in passive recordings and using the material described in Section 4.2.2, four different oddball conditions were designed, in which rising and falling f_0 contours alternated as standard/deviant sounds. In two conditions, standard/deviant sounds are composed of accentual contours (condition1: standard accentual rise/deviant accentual fall; condition2: standard accentual fall/deviant accentual rise), whereas in the other two conditions, standard/deviant sounds are composed of boundary contours (condition3: standard

4 Unravelling the role of intonation in attention orienting using ERPs

boundary rise/deviant boundary fall; condition4: standard boundary fall/deviant boundary rise). While listening to the stimuli, participants watched a nature documentary film with no sound (Deep Blue; Fothergill & Byatt 2003).

4.2.1 Participants

ERP data from 32 right-handed participants were recorded. Handedness was assessed by the Edinburgh Handedness Inventory (Oldfield 1971). All participants were monolingual native speakers of German (28 female, 4 male), aged between 19 and 33 years old (mean age = 24.5 years, $SD = 3.5$). Participants provided written informed consent in accordance with the Declaration of Helsinki and in compliance with the ethics clearance from the Ethics Board of the *Deutsche Gesellschaft für Sprachwissenschaft* (DGfS). Participants received reimbursement for their participation (either course credit or monetary compensation). None of them reported any speech, hearing, or neurological impairment.

4.2.2 Speech materials

The auditory stimuli used in the oddball paradigm comprise rising and falling f_0 contours realised either on the stressed syllable (accentual contours), or at the right boundary, that is, on the last syllable, (boundary contours) of four different lexical items (*Banane* ‘banana’, *Limone* ‘lime’, *Marone* ‘chestnut’, *Melone* ‘watermelon’), resulting in a total of 16 tokens (4 contours x 4 lexical items). For segmental comparability of the lexical items, I selected common trisyllabic German nouns with simple segmental structure and primary lexical stress on the penultimate syllable (CV.CV.CV)³ which were mainly composed of voiced sounds to enable continuous f_0 tracking. Additionally, to control for potential word frequency effects on prominence perception (see Cole et al. 2010), all items were of approximately the same frequency class.⁴

All stimuli were produced by a phonetically trained female native speaker of German and recorded with a sampling rate of 44100 Hz and 16-bit resolution (mono). To ensure natural speech production of the items, the speaker was asked to spontaneously produce all items in isolation, with all intonational contours. To circumvent inconsistencies in the realisation of the same intonation contour across the different items, the most natural-sounding item in terms of speech rate,

³Bold face indicates lexical stress.

⁴The frequency distributions of the words were extracted from the Leipzig Deutscher Wortschatz corpus (Quasthoff & Richter 2005).

duration, pitch range, and meaning⁵ was selected for each contour and presented repeatedly to the speaker as a prompt for the production of the same contour type across lexical items. The speaker repeated the prompt with as little delay as possible, resulting in a natural but very consistent production of the different contours across items, as confirmed by acoustic analysis. The original production of all stimuli was used in the experiment, normalised at -23 LUFS. Figure 4.1 illustrates the mean f_0 contours as well as the individual f_0 contours of all four items superimposed on each other. The similarity of the f_0 contours indicates that they were produced in a very consistent manner across items (see bottom panels).

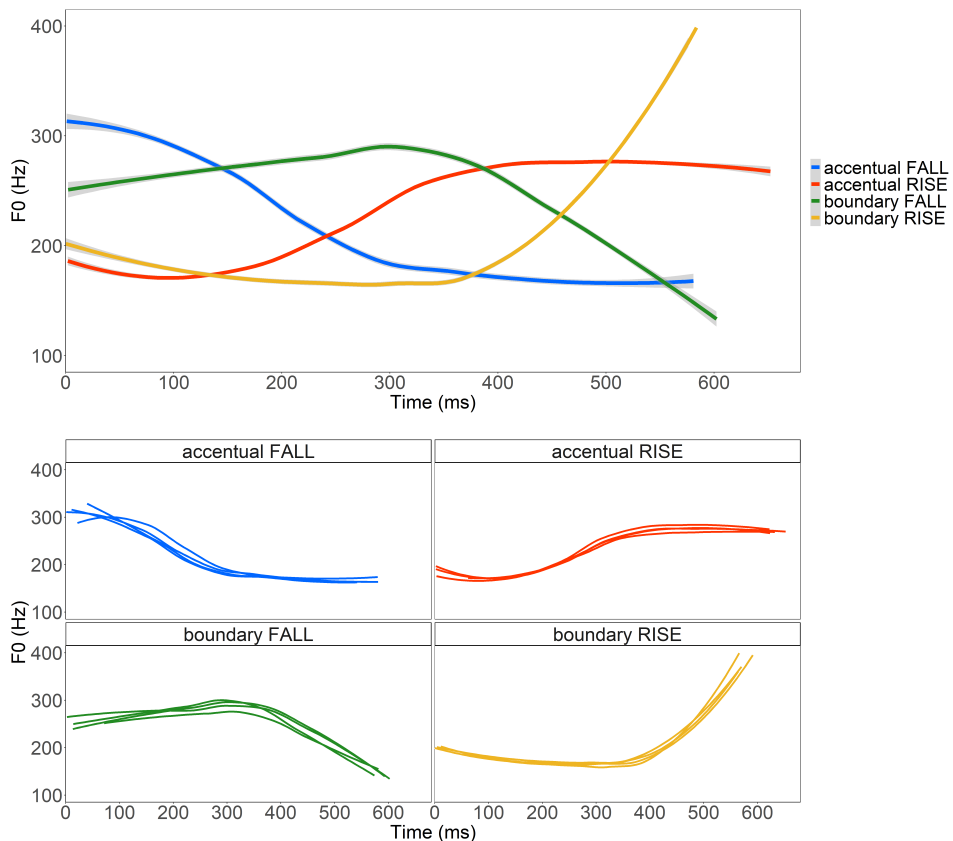


Figure 4.1: Top panel: mean f_0 contours across items. Bottom panel: individual f_0 contours of all items, superimposed for each condition.

⁵That is, a falling contour marking finality, a rising contour marking continuity.

4 Unravelling the role of intonation in attention orienting using ERPs

For the acoustic characterisation of the stimuli, the *ProPer* toolbox was used. ProPer is an open source toolbox for the acoustic analysis of prosodic phenomena based on continuous measurements of periodic energy and f0 (Albert et al. 2018, Cangemi et al. 2019, Albert et al. 2020, Albert 2023). Prior to the ProPer analysis, f0 contours were extracted and corrected manually in Praat (Boersma & Weenink 2024), using a customised version of *mausmooth* (Cangemi 2015). To acoustically describe the current stimuli, the relative periodic energy mass (henceforth, Mass) and the relative Delta f0 (henceforth, $\Delta f0$) metrics were employed. All analyses were conducted on the basis of syllabic units. Scripts and data tables of the current analysis have been made available on the Open Science Framework (OSF) platform (<https://osf.io/nhy4c>).

Mass is the area under the periodic energy curve between two syllabic boundaries. It is calculated as the integral of duration and power, accounting for these two cues together in one variable, capturing the overall prosodic strength of the corresponding syllable (for more on the components of Mass and its improved acoustic characterisation of prominence, see Albert 2023). Here, relative Mass is used, indicating the prosodic strength of one syllable relative to the other syllables in the word. Mass values are calculated as the area under the periodic energy curve of the entire word divided by the number of syllables in the word (Mass = raw Mass of the word/n of syllables). The average Mass is centred around one, so that weak syllables exhibit values lower than one, and strong syllables exhibit values higher than one (weak < 1 < strong).

The measure of $\Delta f0$ describes the f0 trajectory across syllables, using both f0 and periodic energy. First, $\Delta f0$ measures the f0 at the centre of Mass within each syllabic interval, and then the difference in f0 between two subsequent syllables is calculated. Thus, $\Delta f0$ indicates the f0 change from syllable to syllable by calculating the difference from the previous one. For the first syllable, $\Delta f0$ is calculated relative to the speaker's f0 median. The raw $\Delta f0$ is measured in Hz; in this analysis relative $\Delta f0$ values (relative $\Delta f0$ = raw $\Delta f0$ /speaker's f0 range) are used (for more on $\Delta f0$, see Albert 2023).

Figure 4.2 shows a representative example of a *perigram*, a visual representation of the f0 curve, for the item *Melone* realised with a rising pitch accent. The f0 curve is enriched with periodic energy, such that the thicker the line, the greater the periodic energy. The lower part of the figure depicts the periodic energy curve itself. Mass and $\Delta f0$ values are illustrated on the same figure.

The consistent production of the f0 contours across items is also evident in the ProPer analysis. Figure 4.3 depicts relative $\Delta f0$ values per syllable as a function of contour, across items as well as per item. In accentual falling contours (see depiction in blue in the top panels), $\Delta f0$ starts H(igh) on the first syllable, falls

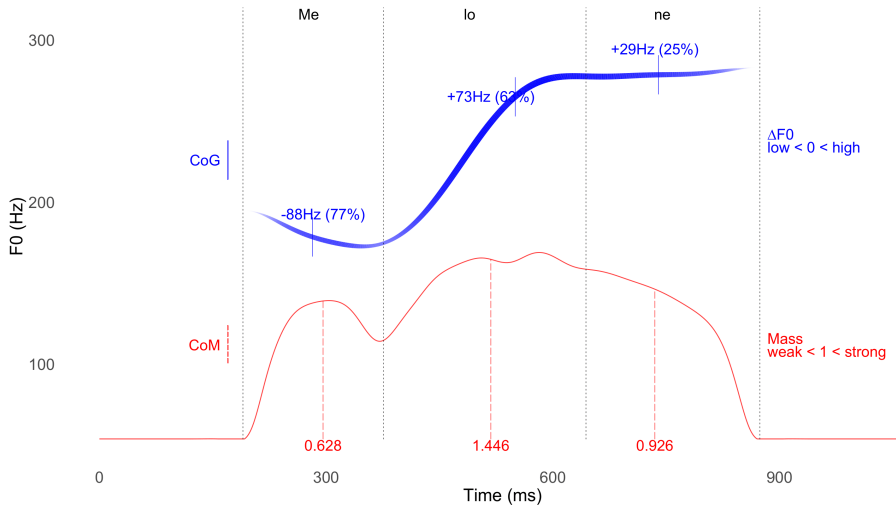


Figure 4.2: A representative example of a periogram. In the upper half of the figure, the blue line illustrates the f_0 curve, with thickness reflecting its strength. The red line in the lower half part of the figure depicts the periodic energy curve. The Hz values on the y-axis correspond only to the f_0 curve. The red vertical dashed lines indicate the position of the centre of Mass (CoM) within syllabic intervals. Mass values on the relative scale are given in numbers below each interval. The short vertical blue lines on the f_0 curve indicate the centre of gravity (CoG), with Δf_0 values displayed on the f_0 curve.

to a L(ow) on the stressed syllable, and levels (or slightly rises) towards the last syllable in the word. In accentual rising contours (see depiction in red in the top panels), Δf_0 starts L on the first syllable, rises to a H on the subsequent stressed syllable, and levels towards the last syllable. In both falling and rising boundary contours (see bottom panels), Δf_0 remains on the same level from the first to the subsequent stressed syllable, and then for the former contour (see depiction in green), falls to a L towards the last syllable (boundary) of the word, while for the latter (see depiction in yellow), it rises to a H.

Moving on to Mass, Figure 4.4 presents Mass values per syllable of each item across accentual and boundary contours. Visual inspection of Mass in both accentual and boundary contours shows a slight variability on the values across items. This is expected considering the different phonetic makeup of each syllable across the different lexical items. Despite these slight differences, the second syllable, which is stressed and accented, exhibits the greatest Mass across items, in the accentual contours, indicating that it is prosodically the strongest syllable in the word, while its overall prosodic strength is similar across rising/falling

4 Unravelling the role of intonation in attention orienting using ERPs

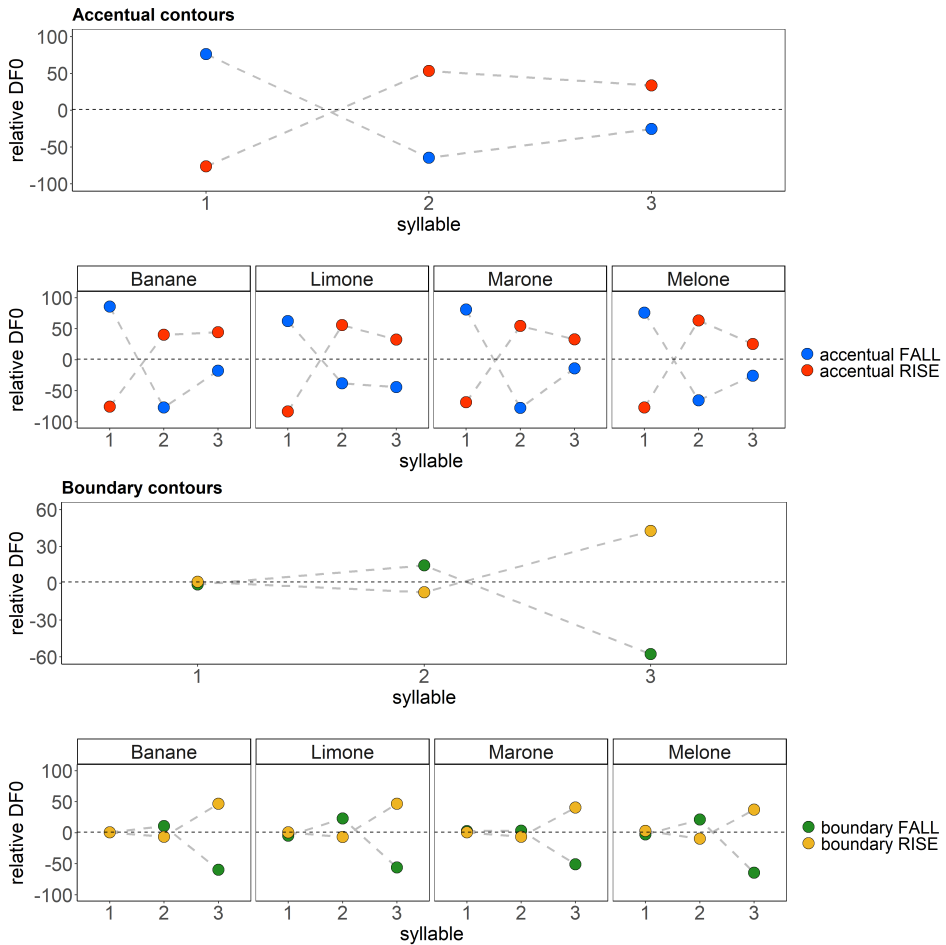


Figure 4.3: Relative Δf_0 values per syllable (x-axis) across accentual contours (top panels; accentual falls in blue, accentual rises in red) and boundary contours (bottom panels; boundary falls in green, boundary rises in yellow) presented across words (first and third panels) as well as separately for each word (second and fourth panels).

contours. In the boundary contours, prosodic strength does not differ greatly between the stressed/accented (second) and the last syllable, which carries the boundary tone. Mass on the stressed syllable is slightly higher in boundary falls than in boundary rises, while Mass on the last syllable is subtly higher in boundary rises than in boundary falls.

Table 4.1 presents means and standard deviations for (i) relative Δf_0 and (ii) relative Mass values per syllable across items for each contour, as well as (iii) the total duration of items.

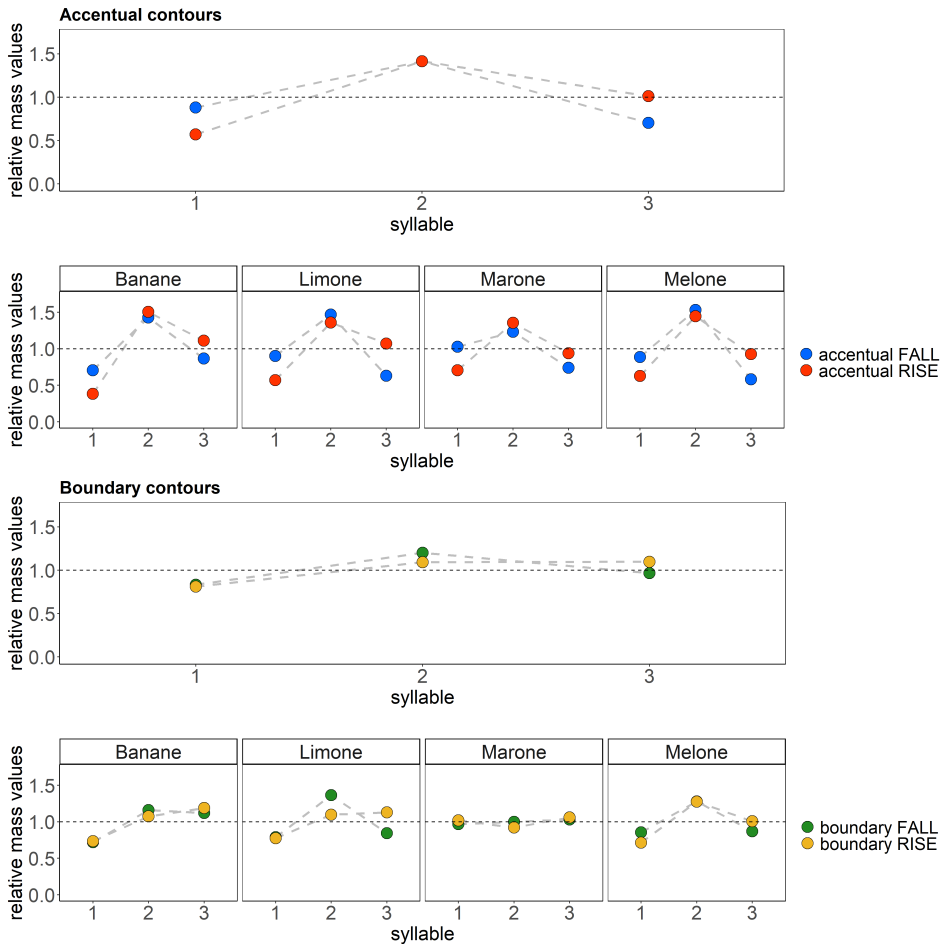


Figure 4.4: Relative Mass values per syllable (x-axis) as a function of item and intonational contour. Top panels illustrate Mass values in accentual contours (rises in red, falls in blue). Bottom panels present Mass values in boundary contours (rises in yellow, falls in green).

4.2.3 EEG recording

The electroencephalogram (EEG) was recorded from 32 Ag/AgCl electrodes, amplified with the Brain Vision amplifier, and digitised at a sampling rate of 1000 Hz. The electrodes were mounted in an elastic EEG cap (EasyCap, EasyCap GmbH, Herrsching, Germany) and placed on the scalp according to the standard international 10–20 system. The electrical contact between scalp and electrodes was achieved by applying a conductible electrolyte. As the MMN is well-documented

4 Unravelling the role of intonation in attention orienting using ERPs

Table 4.1: Mean and standard deviation (in brackets) for relative Δf_0 and Mass values, per syllable, across items for each contour, as well as for the total duration of items.

syllable	accentual fall	accentual rise	boundary fall	boundary rise
<i>relative Δf_0</i>				
1	76.20 (9.99)	-76.13 (6.05)	-1.10 (3.5)	1.20 (1.2)
2	-64.50 (18.5)	53.50 (9.5)	14.50 (9.2)	-7.40 (1.5)
3	-25.50 (13.5)	33.70 (7.8)	-57.80 (5.8)	42.70 (4.7)
<i>relative Mass</i>				
1	0.88 (0.13)	0.57 (0.14)	0.83 (0.10)	0.81 (0.14)
2	1.42 (0.13)	1.42 (0.07)	1.20 (0.16)	1.09 (0.15)
3	0.70 (0.13)	1.01 (0.09)	0.97 (0.13)	1.10 (0.08)
<i>duration [ms]</i>				
item	0.600 (0.01)	0.680 (0.02)	0.630 (0.02)	0.620 (0.02)

to have a frontocentral topography (Näätänen et al. 2007), a distribution of mostly frontal electrodes (AF3/4/7/8, F3/4/7/8, Fz, FC1/2/5/6, FCz, C3/4, Cz, CP1/2, CPz, P3/4/7/8, Pz, POz, Oz) was selected. The AFz electrode position served as the ground electrode, and additional electrodes were placed on the left and right mastoids for referencing (left) and re-referencing (right) of the EEG channels. To control for eye-movement artifacts, the electrooculogram (EOG) was also recorded with electrodes placed to left and right mastoids at the level of the external canthus of each eye, as well as to the supra- and infra-orbital foramina of the right eye. Electrode impedances were kept below 3 k Ω . Figure 4.5 illustrates the distribution of the electrodes used.

4.2.4 Procedure

After electrode application, participants were seated in a booth with sound insulation on a comfortable chair, in front of a monitor. As there was no active task for this experiment, participants were asked to relax and watch a nature documentary film with no sound (Deep Blue; Fothergill & Byatt 2003). Participants were informed that during the film they would hear some audio unrelated to the film over the loudspeakers, but they were asked to ignore it.

Participants were presented with standard/deviant f_0 contours realised on the same item within condition, but the items across conditions always differed

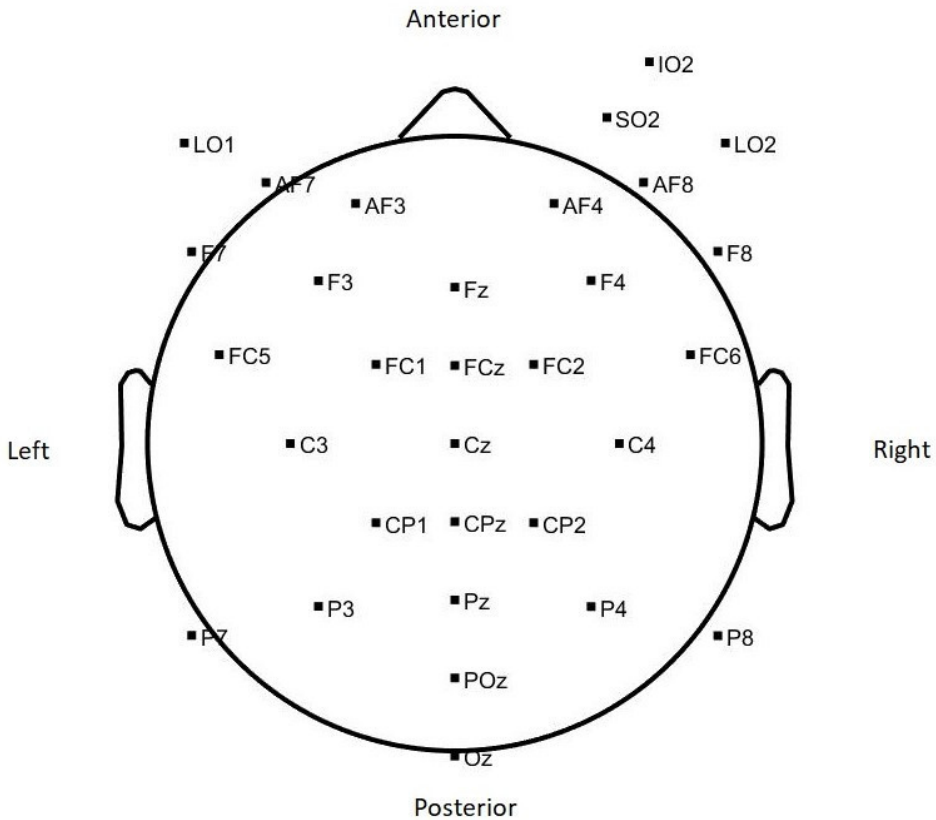


Figure 4.5: 2D layout of electrode placement according to the 10–20 system. In this system, each electrode is named by one or two letters to indicate the general brain region (Fp = frontal pole, F = frontal, C = central, P = parietal, O = occipital, T = temporal). Each letter is followed by a number representing the hemisphere (odd numbers = left hemisphere, even numbers = right hemisphere). The lowercase z stands for the number zero, which represents the electrodes on the midline. The higher the number, the greater the distance from the midline.

(Latin Square Design: 4 items x 4 oddball conditions). To control for systematic order and frequency effects potentially induced by the exposure to the oddball condition and/or item order, 16 fully-counterbalanced lists were created with different oddball condition order and item, so that each list presented all items and oddball conditions but never the same item across conditions and never the same condition order. Each participant heard only one of the lists (the exact distribution of the lists have been made available on OSF [<https://osf.io/nhy4c/>]).

4 Unravelling the role of intonation in attention orienting using ERPs

Each oddball condition consisted of 1000 trials in total (850 standards and 150 deviants; the inter-stimulus interval was jittered between 450–545 ms in order to achieve the same stimulus onset asynchrony for all items, which was 1147 ms), resulting in a presentation time of approximately 20 min per condition and approximately 1 hour and 20 min in total. The order of the trials was fully randomised, with at least two consecutive standards between deviants in order to avoid deviant stimuli developing their own memory trace. The initial 15 standards of each condition were excluded from all subsequent analyses, as their presentation served the sensory-memory trace formation (e.g., Näätänen et al. 2019), resulting in a total of 3940 trials (985 trials x 4 conditions). Figure 4.6 depicts a schematic illustration of the oddball conditions.

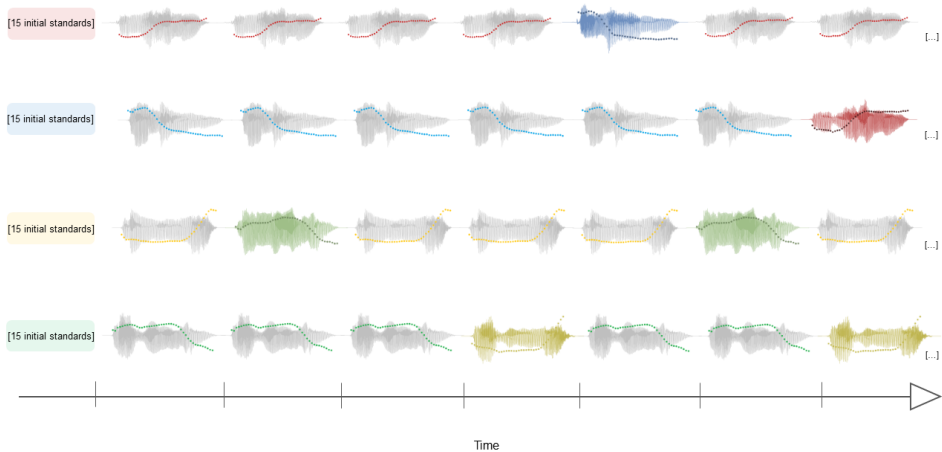


Figure 4.6: Schematic illustration of the oddball conditions. The four rows present the four different oddball conditions. From top to bottom, the first two rows illustrate the oddball conditions in which accentual rising (in red) and accentual falling (in blue) intonation alternate as standard/deviant within condition; the last two rows depict the boundary rising (in yellow) and boundary falling (in green) oddball alternation within conditions. Each condition started with the presentation of 15 standard sounds to serve the creation of the memory trace.

4.2.5 Data pre-processing

The data were pre-processed using the Matlab-based toolbox EEGLAB (Delorme & Makeig 2004), which was developed at the Swartz Center for Computational Neuroscience. To reduce computational demands, the first step was to resample

the data to 250 Hz. Afterwards, the data were re-referenced to linked mastoids. Next, an independent component analysis (hereafter, ICA) for artifact correction was performed. For ICA decomposition, the EEG was filtered with a 1 Hz high-pass filter to approach stationarity, and a 45 Hz low-pass filter to remove line noise. Subsequently, artifact components (muscle and eye components above 80%; heart components above 90%) were automatically detected and removed from the raw EEG data. After artifact rejection, the raw EEG data were filtered with a 0.3 Hz high-pass and a 30 Hz low-pass filter, instead of baseline correction (see Friederici et al. 2000, Wolff et al. 2008, Widmann et al. 2015, Maess et al. 2016a,b). Thereafter, the data were epoched from -200 to 1000 ms post stimulus onset. For reproducibility, the pre-processing script can be found on OSF (<https://osf.io/nhy4c/>).

4.2.6 Post-processing and statistical data analysis

Post-processing and statistical analyses were conducted in R, version 4.1.2 (R Core Team 2023). For data processing and visualisation, the R packages *tidyverse* 1.3.1 (Wickham et al. 2019), and *ggplot2* 3.3.5 (Wickham 2016) were used. ERP amplitude was analysed by fitting Bayesian hierarchical regression models using the *brms* 2.17.0 package (Bürkner et al. 2023). Data and scripts can be found on OSF.

4.2.6.1 Post-processing

After data epoching, to avoid effects of the repeated number of standards (recall that the deviant trials formed only 15% of total trials), an equal number of standard and deviant trials entered the statistical analyses. To achieve this, only the standards that appeared directly before a deviant were selected, yielding a total of 300 trials (150 standards/150 deviants) per electrode site.

The current analysis focuses on two ERP effects that are claimed to index activation of pre-attentive and conscious attentional mechanisms, respectively: the MMN and the P3 responses (see Section 2.5). The MMN is a negative auditory ERP component which is traditionally obtained as a difference wave by subtracting the ERPs to standard from those to deviant stimuli (i.e., deviant ERPs – standard ERPs; Näätänen et al. 2019). However, this approach requires the use of the grand averaged signal, leading to a great loss of variance in the data. In the current analysis, ERP amplitude was averaged by time window for every participant, electrode site, and trial. This allows for fitting the models on single-trial data and at the same time model variance associated to each participant.

4 Unravelling the role of intonation in attention orienting using ERPs

As discussed in Section 2.5, MMN is reported to typically peak between 100 and 250 ms after stimulus onset (e.g., Duncan et al. 2009). MMN is usually followed by a positive ERP component, the P3 response, around 300 ms or later after stimulus onset (P3 latency depends on the complexity of the processing: the more complex the processing, the longer the latency, varying approximately from 250 to 1000 ms; see Duncan et al. 2009). Nevertheless, there is considerable variability in the definition of the time windows that have been used to analyse these effects, as peak latency has been usually defined on the basis of visual inspection of the difference waves. For example, there have been studies defining time windows with MMN peak latency at around 350 ms (e.g., Emmendorfer et al. 2020). For this reason, ERP amplitude is here analysed from 0 to 700 ms after stimulus onset in steps of 100 ms, resulting in seven time windows (0–100 ms, 100–200 ms, 200–300 ms, 300–400 ms, 400–500 ms, 500–600 ms, 600–700 ms). Further, as the MMN is well-documented to have a frontocentral topography, and the P3 a frontal distribution (e.g., Duncan et al. 2009), for the current analyses a spatial region of interest⁶ was defined consisting of the AF3, AF4, F3, Fz, F4, FC1, FCz, FC2, and Cz electrode sites.

4.2.6.2 Inference criteria

ERP amplitude (in microvolt) was modelled from 0 to 700 ms after stimulus onset by fitting separate Bayesian hierarchical regression models per oddball condition, in steps of 100 ms. Treatment contrast was used to code the predictor SOUND (levels; standard/deviant) with the level *standard* serving as the reference level. Random effects for SUBJECTS included full variance–covariance matrices (e.g., Barr et al. 2013). Weakly informative priors were used for all the parameters (the full prior specification can be found in the script provided on OSF), as they allow for a wide range of effect sizes but control for unreasonably large effects. All models ran with four chains and 4000 iterations with a warm-up period of 2000 iterations. There were no divergent transitions and all \hat{R} s were close to 1, showing that chains mixed without issues. Model fits were also visually inspected using the posterior predictive check function.

In the following section, inferences are drawn from the posterior distributions of the parameters. For this, posterior estimates, the low and high boundaries of

⁶At first, I intended to run the models using all registered electrodes including sagittality and laterality as predictors in the models. However, besides the theoretical reasons I report here, there were also practical reasons that led to the decision of defining one spatial region of interest – the models including all scalp electrode sites were highly computationally demanding, to the point of being intractable.

the 90% credible interval (CrI) of the estimate, and the posterior probability that the estimate falls on one side of zero (e.g., $P(\beta < 0) = 0.95$) are reported. When almost all of the posterior mass for an estimate lies on one side of zero, zero is not included in the 90% CrI (by a reasonably clear margin), and the posterior probability P is close to one, the effect is considered reliable.

4.3 Results

Figure 4.7 illustrates the grand averaged ERP waves per oddball condition time-locked to the onset of stimulus, as depicted by the vertical dashed line (see also Figure 4.8 for the grand averaged difference waves obtained by subtracting standards from deviants per oddball condition). Left panels show ERPs to the two accentual oddball conditions in which accentual falls (in blue) and rises (in red) alternate as standard/deviant sounds. Right panels present ERPs to the two boundary oddball conditions in which boundary falls (in green) and rises (in yellow) alternate as standard/deviant sounds. All panels depict ERPs to standards in black, and ERPs to deviants in colour. Visual inspection of the waves reveals that all contour types, when presented as deviants, evoked an MMN activity relative to their corresponding standard stimulation with an onset around 200 ms. For all deviants except the accentual fall (left top panel), the MMN activity (coloured area between ERPs to standards and ERPs to deviants) appears to last for two successive time windows (200–400 ms), with the accentual rising deviant (left bottom panel) exhibiting the most pronounced effect. MMN to falling deviants (both accentual and boundary; top panels) appears to be followed by an additional P3, at the 400–500 ms time window (shaded area between waves) for the accentual falls, and at a later time window (500–600 ms) for boundary falls.

In what follows, I will first report results on the difference between deviants and standards (i.e., standard sound vs. deviant sound) per oddball condition to detect whether MMN and P3 responses were elicited by deviant sounds relative to their standard stimulation. A brain activity is referred to as an MMN when a reliable negative difference between deviant and standard sounds is detected during the 100–200 ms, 200–300 ms, and 300–400 ms time windows. A brain activity is identified as a P3 response when a reliable positive difference between deviant and standard stimulation is observed in the 300–400 ms window or at a later time window. Subsequently, I will summarise MMN and/or P3 effects (if any) in the presence of rises vs. falls, regardless of their position (accentual/boundary), aiming to establish whether the direction of the pitch movement affected the evoked brain response. Lastly, I will sum up MMN and/or P3 effects in the light

4 Unravelling the role of intonation in attention orienting using ERPs

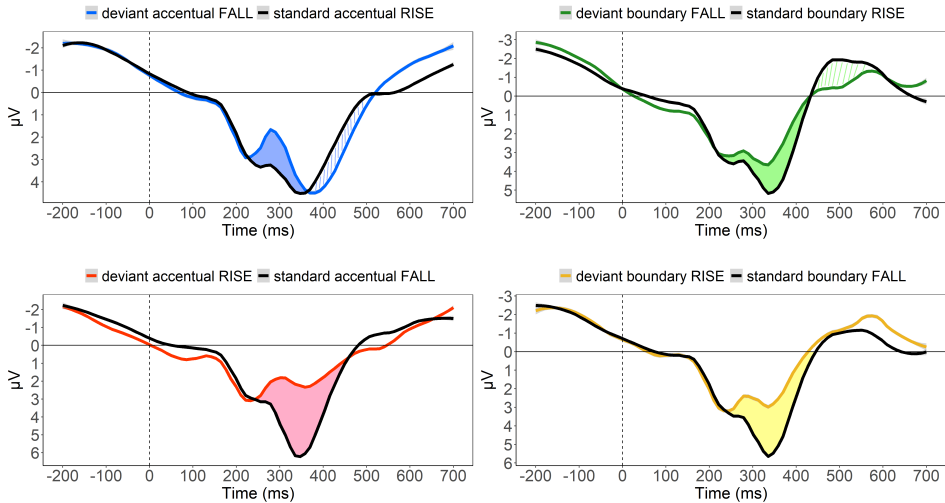


Figure 4.7: Grand averaged ERP waves (per oddball sequence) recorded to the onset of stimulus (illustrated by the vertical dashed line) over time (x-axis) at the AF3, AF4, F3, Fz, F4, FC1, FCz, FC2, and Cz electrode sites. Negative voltage is plotted upwards. Left panels show ERPs to the two accentual oddball conditions in which accentual falls (in blue) and rises (in red) alternate as standard/deviant sounds. Right panels present ERPs to the two boundary oddball conditions in which boundary falls (in green) and rises (in yellow) alternate as standard/deviant sounds.

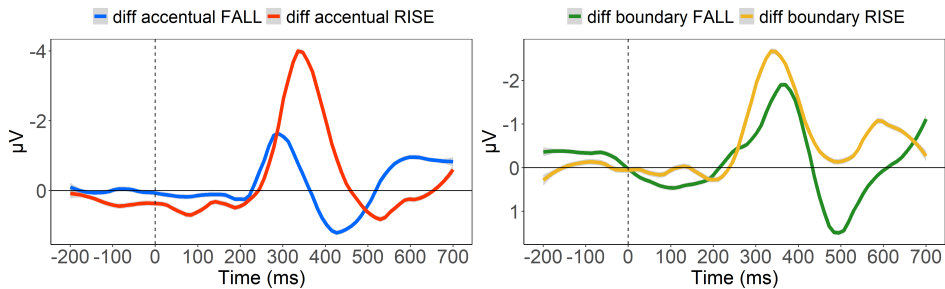


Figure 4.8: Grand averaged difference waves time-locked to the onset of stimulus (illustrated by the vertical dashed line) and obtained in AF3, AF4, F3, Fz, F4, FC1, FCz, FC2, and Cz electrode sites. Negative voltage is plotted upwards. Difference waves were obtained by subtracting standard waves from deviant waves per oddball condition. Left side panels depict difference waves of accentual falling deviants (in blue) and accentual rising deviants (in red). Right side panels show difference waves of boundary falling deviants (in green) and boundary rising deviants (in yellow).

of the position of the rise and fall to find whether the phonological association (i.e., status) of the tonal event affects the elicitation of MMN and/or P3 responses.

4.3.1 Accentual contours

Figure 4.9 illustrates the posterior distributions of the estimated effects for the difference between standard and deviant sounds in the two accentual oddball conditions per time window, in ascending order. The estimated differences between accentual falling deviants and accentual rising standards (oddball condition 1) are depicted in blue; the estimated differences between accentual rising deviants and accentual falling standards (oddball condition 2) are depicted in red.

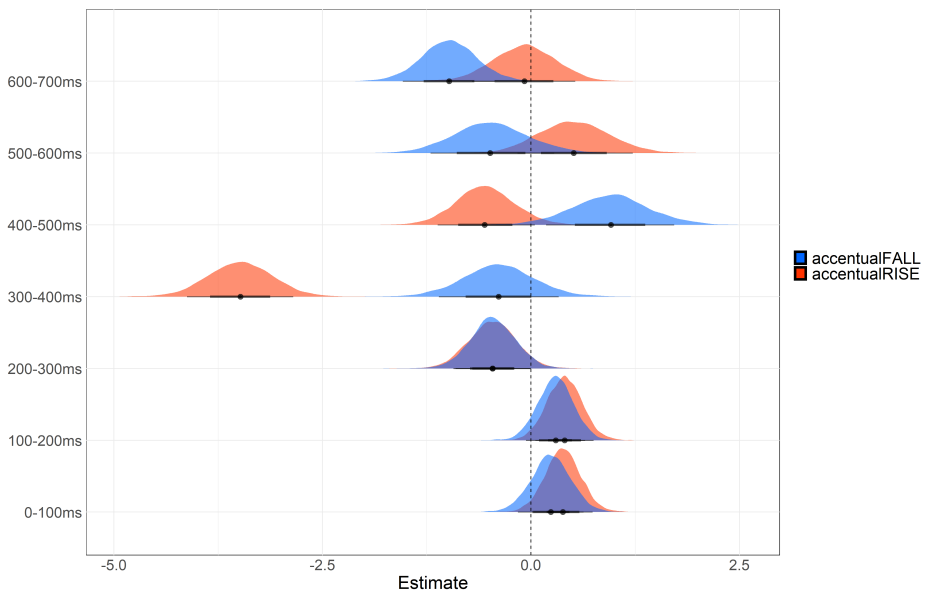


Figure 4.9: Posterior distributions of the estimated effects for the difference between standard and deviant sounds in the two accentual oddball conditions. Accentual falling deviants vs. accentual rising standards are depicted in blue. Accentual rising deviants vs. accentual falling standards are shown in red. Time windows are presented in ascending order. Error bars around the posterior means represent 66% (thick) and 90% CrI.

4.3.1.1 Accentual falling deviant vs. accentual rising standard

Table 4.2 presents model details on ERP amplitude differences between deviant accentual falling contours and standard accentual rising contours per time win-

4 Unravelling the role of intonation in attention orienting using ERPs

dow. In short, the model revealed no reliable differences between deviant accentual falling and standard accentual rising contours in the 0–100 ms, 100–200 ms, and 500–600 ms time windows. By contrast, between 200 and 300 ms, the model very strongly favours the interpretation of a negative-going difference in amplitude, indicating an MMN activity, although the 90% CrI includes zero on the margin. The model also estimated a reliable positive difference between 400 and 500 ms, indicating that the MMN to falling deviants was followed by an additional P3, and a further negative-going difference between 600 and 700 ms.

Table 4.2: Tabular overview of modelling results for accentual falling deviants vs. accentual rising standards per time window.

Time Window	β	SE	Low CrI	High CrI	Evid. Ratio	Post. Prob.
0–100 ms	0.24	0.24	-0.16	0.63	5.46	$P(\beta > 0) = 0.85$
100–200 ms	0.30	0.21	-0.06	0.65	10.89	$P(\beta > 0) = 0.92$
200–300 ms	-0.46	0.27	-0.90	0.00	19.78	$P(\beta < 0) = 0.95$
300–400 ms	-0.39	0.43	-1.10	0.33	4.61	$P(\beta < 0) = 0.82$
400–500 ms	0.95	0.46	0.18	1.72	44.20	$P(\beta > 0) = 0.98$
500–600 ms	-0.48	0.45	-1.20	0.28	6.04	$P(\beta < 0) = 0.82$
600–700 ms	-0.98	0.33	-1.54	-0.43	532.33	$P(\beta < 0) = 1.00$

4.3.1.2 Accentual rising deviant vs. accentual falling standard

Table 4.3 outlines model details for ERP amplitude contrasts between accentual rising deviants and accentual falling standards per time window. For these contrasts, the model estimated a reliable positive difference in amplitude in the 0 to 100 ms and 100 to 200 ms time windows. For the next two successive time windows, 200–300 ms and 300–400 ms, compelling evidence for a negative difference in amplitude was found, suggesting the presence of MMN activity. For the remaining time windows, the model did not suggest any reliable amplitude differences.

Overall, these results show that both accentual rises and falls, when presented as deviant sounds, elicited an MMN activity, starting in the 200 to 300 ms time window. For the accentual rising deviants, the MMN activity appears to have a longer duration than it does for the accentual falling deviants, as in this condition a negative difference is present for two successive time windows (i.e., from 200 to 400 ms). For the accentual falling deviants, the results show that their

Table 4.3: Tabular overview of modelling results for accentual rising deviants vs. accentual falling standards per time window.

Time Window	β	SE	Low CrI	High CrI	Evid. Ratio	Post. Prob.
0–100 ms	0.38	0.22	0.03	0.74	25.58	$P(\beta > 0) = 0.96$
100–200 ms	0.40	0.21	0.06	0.75	34.56	$P(\beta > 0) = 0.97$
200–300 ms	-0.46	0.29	-0.93	0.00	18.51	$P(\beta < 0) = 0.95$
300–400 ms	-3.49	0.39	-4.13	-2.83	Inf	$P(\beta < 0) = 1.00$
400–500 ms	-0.55	0.36	-1.12	0.05	14.78	$P(\beta < 0) = 0.94$
500–600 ms	0.52	0.42	-0.16	1.22	8.22	$P(\beta > 0) = 0.89$
600–700 ms	-0.08	0.38	-0.71	-0.53	1.42	$P(\beta < 0) = 0.59$

MMN activity is followed by an additional P3 response in the 400 to 500 ms time window, yet accentual rising deviants did not engender such a brain response.

4.3.2 Boundary contours

Figure 4.10 presents posterior distributions of the estimated effects for the differences between standard and deviant sounds in the two boundary oddball conditions per time window, in ascending order. The estimated differences between boundary falling deviants and boundary rising standards (oddball condition 3) are depicted in green; the estimated differences between boundary rising deviants and boundary falling standards (oddball condition 4) are depicted in yellow.

4.3.2.1 Boundary falling deviant vs. boundary rising standard

Model details on the ERP amplitude differences between boundary falling deviant and boundary rising standard contours per time window are presented in Table 4.4. Briefly, the model did not provide reliable differences in the 0 to 100 ms, 100 to 200 ms, 400 to 500 ms, and 600 to 700 ms time windows, but showed compelling evidence for a negative difference in amplitude between 200–300 ms and 300–400 ms, indexing the elicitation of an MMN activity. Although the model estimated a positive difference between 400 and 500 ms, this difference was not reliable. Nonetheless, the model provided compelling evidence for another positive-going difference in the 500 to 600 ms time window, suggesting the presence of a P3 response.

4 Unravelling the role of intonation in attention orienting using ERPs

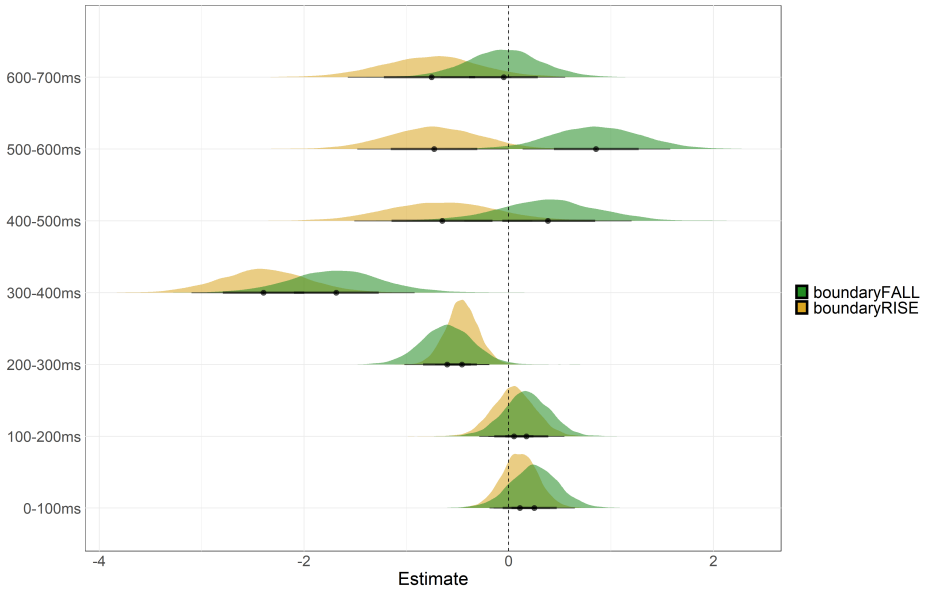


Figure 4.10: Posterior distributions of the estimated effects for the differences between standard and deviant sounds in the two boundary oddball conditions. Boundary falling deviants vs. boundary rising standards are depicted in green. Boundary rising deviants vs. boundary falling standards are shown in yellow. Time windows are presented in ascending order. Error bars around the posterior means represent 66% (thick) and 90% CrI.

Table 4.4: Tabular overview of modelling results for boundary falling deviants vs. boundary rising standards per time window.

Time Window	β	SE	Low CrI	High CrI	Evid. Ratio	Post. Prob
0–100 ms	0.25	0.24	-0.15	0.65	5.93	$P(\beta > 0) = 0.86$
100–200 ms	0.17	0.23	-0.20	0.54	3.68	$P(\beta > 0) = 0.79$
200–300 ms	-0.60	0.25	-1.02	-0.19	107.11	$P(\beta < 0) = 0.99$
300–400 ms	-1.68	0.46	-2.43	-0.92	1599	$P(\beta < 0) = 1.00$
400–500 ms	0.39	0.49	-0.43	1.20	3.82	$P(\beta > 0) = 0.79$
500–600 ms	0.85	0.44	0.14	1.58	34.4	$P(\beta > 0) = 0.97$
600–700 ms	-0.05	0.37	-0.66	0.55	1.24	$P(\beta < 0) = 0.55$

4.3.2.2 Boundary rising deviant vs. boundary falling standard

Model details on the ERP contrasts between boundary rising deviants and boundary falling standards per time window are shown in Table 4.5. No evidence was found for a difference in the 0 to 100 ms, 100 to 200 ms, and 400 to 500 ms time windows, but the model provided compelling evidence for a negative difference in amplitude for the 200 to 300 ms and 300 to 400 ms time windows, indicating an MMN activity. For the last two time windows (500–600 ms and 600–700 ms), the model estimated a negative-going difference. Here, although the 90% CrI includes zero on the margin, it could still favour the interpretation of a (marginal) difference.

Table 4.5: Tabular overview of modelling results for boundary rising deviants vs. boundary falling standards per time window.

Time Window	β	SE	Low CrI	High CrI	Evid. Ratio	Post. Prob
0–100 ms	0.11	0.18	-0.19	0.40	2.75	$P(\beta > 0) = 0.73$
100–200 ms	0.05	0.21	-0.29	0.39	1.49	$P(\beta > 0) = 0.60$
200–300 ms	-0.46	0.16	-0.72	-0.19	379.95	$P(\beta < 0) = 1.00$
300–400 ms	-2.39	0.43	-3.10	-1.68	Inf	$P(\beta < 0) = 1.00$
400–500 ms	-0.65	0.53	-1.51	0.23	8.36	$P(\beta < 0) = 0.89$
500–600 ms	-0.73	0.45	-1.48	0.00	18.75	$P(\beta < 0) = 0.95$
600–700 ms	-0.77	0.48	-1.57	0.03	17.43	$P(\beta < 0) = 0.95$

I have shown that rising and falling boundary contours highlight an MMN activity relative to their corresponding standard stimulation, lasting from 200 to 400 ms. Furthermore, brain responses to boundary rises and falls differ in that the MMN to boundary falls is followed by a P3 response, while there is no evidence for boundary rises eliciting a subsequent positivity.

4.3.3 Interim summary

Figure 4.11 summarises all posterior distributions of the estimated differences per oddball condition in the time windows of interest (i.e., MMN and P3 time windows). Table 4.6 presents an overview of the estimated differences between rising and falling contours as well as between accentual and boundary contours. These differences point towards a distinct relevance of contour direction (i.e., rise or fall) and linguistic context for the attentional mechanisms.

4 Unravelling the role of intonation in attention orienting using ERPs

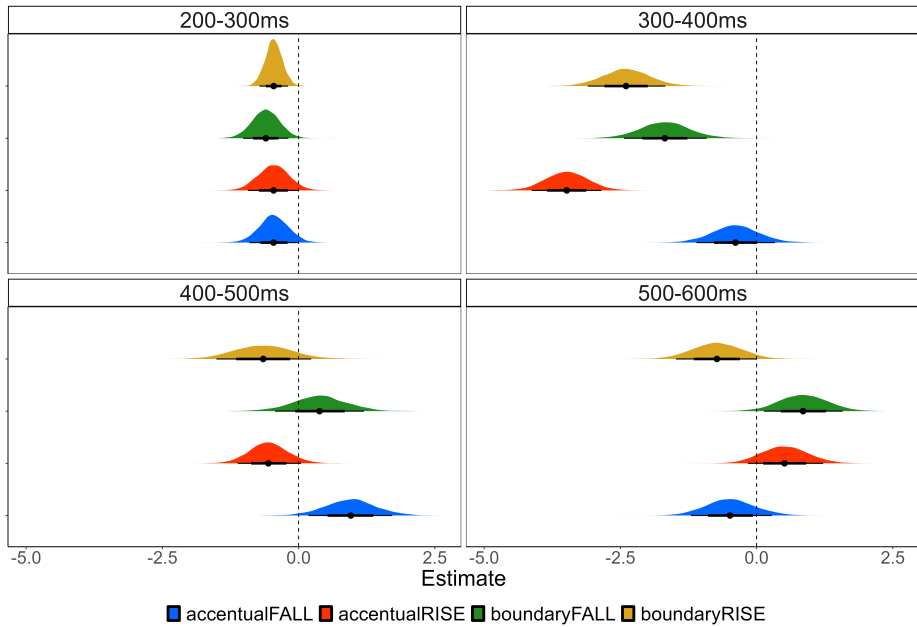


Figure 4.11: Posterior distributions of the estimated effects for the differences between standard and deviant sounds across oddball conditions in the time windows of interest (MMN time windows: 200–300 ms and 300–400 ms; P3 time windows: 400–500 ms and 500–600 ms). From top to bottom, boundary rising deviants vs. boundary falling standards are depicted in yellow, boundary falling deviants vs. boundary rising standards are depicted in green, accentual rising deviants vs. accentual falling standards are depicted in red, and accentual falling deviants vs. accentual rising standards are depicted in blue. Error bars around the posterior means represent 66% (thick) and 90% CrI.

Comparing accentual rising to falling deviants, it was found that both contours elicit an MMN activity with an onset in the 200 to 300 ms time window. Whereas there is no quantitative difference at the beginning of the MMN elicitation between accentual rises and falls, the MMN to accentual rises is prolonged over the subsequent time window (300–400 ms), reflecting a longer-lasting effect compared to the accentual falls. What is more, accentual falls after MMN elicitation engender a subsequent P3 in the 400 to 500 ms time window, a response that is not evoked by accentual rises.

Let me now turn to the rising vs. falling comparison in the boundary conditions. The findings here are quite similar. Both boundary rises and falls evoke an MMN activity, lasting for two successive time windows (200–400 ms), with a tendency for the boundary rises to show a stronger negative effect in the 300–400 ms time window. The evidence from the model is not reliable enough to

Table 4.6: Tabular overview of estimated differences between rising and falling contours as well as accentual and boundary contours in the time windows of interest.

Time Window	β	SE	Low CrI	High CrI	Evid. Ratio	Post. Prob
<i>Rises vs falls</i>						
<i>Accentual</i>						
MMN (200–300 ms)	0.00	0.39	-0.65	0.64	1.02	$P(\beta < 0) = 0.50$
MMN (300–400 ms)	-3.10	0.59	-4.07	-2.15	Inf	$P(\beta < 0) = 1.00$
P3 (400–500 ms)	1.50	0.59	0.54	2.47	132.33	$P(\beta > 0) = 0.99$
<i>Rises vs falls</i>						
<i>Boundary</i>						
MMN (200–300 ms)	0.14	0.30	-0.35	0.64	0.46	$P(\beta < 0) = 0.31$
MMN (300–400 ms)	-0.71	0.62	-1.74	-0.28	7.11	$P(\beta < 0) = 0.88$
P3 (500–600 ms)	1.58	0.63	0.56	2.61	144.45	$P(\beta > 0) = 0.99$
<i>Accent vs Boundary</i>						
<i>Rises</i>						
MMN (200–300 ms)	0.00	0.33	-0.56	0.53	1.01	$P(\beta < 0) = 0.50$
MMN (300–400 ms)	-1.10	0.58	-2.05	-0.15	36.21	$P(\beta < 0) = 0.97$
<i>Accent vs Boundary</i>						
<i>Falls</i>						
MMN (200–300 ms)	-0.14	0.37	0.37	0.46	1.89	$P(\beta < 0) = 0.65$
MMN (300–400 ms)	-1.29	0.63	-2.36	-0.25	53.79	$P(\beta < 0) = 0.98$

claim with confidence that the MMN to boundary rises is more pronounced than the MMN to boundary falls. However, the MMN to boundary falls, similarly to the MMN to accentual falls, is followed by an additional P3. There is no such evidence for boundary rises.

Considering the phonological status of the rise and fall, namely their position in the prosodic structure (accent/boundary), the MMN to falling contours is of a similar magnitude in the starting time window (200–300 ms), regardless of phonological status, while the MMN to boundary falls is prolonged over the subsequent time window (300–400 ms) compared to the activity evoked by the accentual falls. Moreover, the MMN to both falling deviants is followed by an

additional P3, although at different time windows, at the 400 to 500 ms window for accentual falls, and at a later time window (500–600 ms) for boundary falls. Zooming in on rising contours, although the onset of the MMN activity does not differ as a function of the phonological status of the rise (accentual/boundary), accentual rises exhibit the most pronounced effect in the 300 to 400 ms time window. Finally, no evidence is found for a P3 brain response in the presence of either accentual or boundary rises.

I return to these findings in Section 4.4, where I will argue that the observed brain responses in relation to the tested contours could indicate different neurocognitive processes for intonational rises and falls, due to both speech sound complexity and linguistic or context-specific interpretation.

4.4 Discussion

Unlike previous neurophysiological research on unexpected auditory changes in a stream of repetitive stimulation, which focused primarily on non-linguistic stimuli, the current work explored the neural responses to rising and falling pitch attributable to accents and boundary tones of sequentially presented lexical items in German. The main aim of this study was twofold. Using ERP data in an oddball paradigm, this study investigated the relevance of intonational rises for attention orienting, as well as to what extent the phonological status of the rise plays a role. It is important to note that I discuss only MMN- and P3-related effects, as these are the brain responses that relate to my research questions. In this discussion, I will first put the findings in the broader context of rising vs. falling intonation, irrespectively of the prosodic structure, and their relevance for attention orienting (Hypothesis 1). Subsequently, I will discuss whether the phonological status of the rise and the fall (accent vs. boundary) modulates attentional resources differently (Hypothesis 2). Finally, I will briefly discuss some implications of the current findings for the neural architecture of speech perception.

4.4.1 The processing of rises and falls and their contribution to attention orienting

The obtained results provide evidence for distinct neurocognitive processes for intonational rises and falls in the context of an oddball list. This is reflected in the different brain responses observed during the online processing of the pitch contours: rising pitch contours engendered MMN activity, while falling pitch contours evoked an MMN-P3 complex. Whilst MMN indexes automatic

(i.e., pre-attentive) processes related to involuntary attention switch towards the deviant pitch contour, the MMN-P3 complex indicates that processes at the pre-attentive level subsequently activate processes at the conscious level, bringing the perception of the deviant contour into awareness and to voluntary attention.

These processing patterns might reflect the presence of different mechanisms or routes for signal-driven (i.e., bottom-up) and context-driven (i.e., top-down) processes. Although the former have been the target of previous research utilising the oddball paradigm, the presentation of lexical items in the present study has the potential to give rise to contextual expectations for a particular list intonation. The results of this study indicate that, in a linguistic context, the acoustic signal is not the only source for attention orienting, as would be expected in a “meaningless” context of pure sine waves. Pitch, like any other acoustic property, is processed as the sensory input unfolds, while at the same time, expectations for the forthcoming input are formed incrementally (e.g., Röhr et al. 2021). In a linguistic context, the generation of expectations can be based both on the sensory information of the input (signal-driven) and on contextual meaning (context-driven; e.g., Röhr et al. 2021). It is worth elaborating on how the present results elucidate this last point.

In two of the oddball conditions, falling pitch contours were presented as deviant sounds, with their corresponding standard stimulation consisting of repetitive rises. In such sequences, the listener can potentially build predictions derived from two different sources. First, the auditory processing system is able to predict prospective sounds by detecting regularities in the sensory input (signal-driven expectations; e.g., Näätänen 1990, Sussman & Winkler 2001, Friston 2010, Vachon et al. 2012). Thus, the listener, after being exposed to a repetitive sequence of stimuli with rising pitch, predicts and anticipates that the next auditory event will again be a rising contour. Predictions can also arise from the linguistic interpretation of the context (context-driven expectations; for predictions in language, see, among others, Schumacher et al. 2015, Röhr et al. 2021). The oddball paradigm resembles a list context because it presents (auditory) events repeatedly and sequentially. On top of the sequential presentation of the stimuli in this study, naturalistic pitch contours realised on real words were used, approximating an even more natural list context.⁷ List intonation in German, as mentioned in Section 3.5, typically involves rising pitch on non-final and penultimate items (indicating continuity) followed by a fall on the final item (indicating finality; e.g., Baumann & Trouvain 2001, Peters 2018). Thus, the repetitive rising stimulation, as a natural and appropriate contour on non-final list items, denotes that the list

⁷The words all refer to food items that could conceivably be used in, e.g., a shopping list.

is not over yet. The repetitive/standard rises, therefore, might have elicited additional predictions driven by the contextually created meaning. Such an expectation is the anticipation that the list will come to an end at some point. Therefore, the listener, first, anticipates a rising contour on the basis of the sensory information that is available in the repetitive auditory stimulation (signal-driven expectation) and, second, given the available contextual meaning, expects that at some point the list will be over, thus anticipating a falling contour (context-driven expectation). Recall now that when the deviant fall was presented, an MMN-P3 complex was elicited, indicating that, first, the violation of the anticipated rising contour activated a pre-attentive response of an involuntary attention switch to the unexpected falling contour. Subsequently, a conscious or voluntary attention orientation towards this falling contour was observed, potentially induced by the validation of the context-driven expectation, that is, the anticipation that the list would at some point come to an end. Similar findings are reported in a study by Liu et al. (2016) on the neural processing of tone and intonation in Mandarin Chinese. Crucially, the authors observed that P3 is modulated by the context (question vs. statement) for the falling contour Tone4. The present findings are also in accord with Röhr et al. (2021) who showed, in an ERP study, that signal- and context-driven cues consume attentional resources at different processing stages: signal-driven cues attract attention at an early processing stage, while context-driven cues attract attention at a later processing stage (for more on these studies, see Section 3.4).

Now let us move to the other two oddball conditions in this study, where the standard stimulation consisted of repetitive falling pitch, occasionally interspersed with rising pitch. Such sequences give the feeling of the presentation of isolated events, as the repetitive/standard falling intonation is contextually an inappropriate/unnatural pitch contour for non-final items of a sequence. Hence, such sequences might allow the listener to only build signal-driven expectations. Put differently, based on the sensory input, the listener anticipates that the next sound will be a falling contour again, but cannot generate a prediction over and above the purely signal-based one, as this repetitive signal is already unexpected in the context of a list (i.e., incongruency in the prosodic realisation on non-final items in the list context). Recall now that the presentation of the deviant rises evoked an MMN activity, showing that the prediction generated from the sensory input was violated, activating an automatic involuntary attention switch to the unexpected rising contour. Brain responses related to conscious/voluntary attentional processes, such as the P3, were not observed for rising deviants.

To better understand how these distinct processes for the rising and falling deviants, as reflected in the observed brain responses, link to attention orient-

ing and prominence-cueing, a joint consideration of these findings is required. The starting observation is that all pitch contours, when presented as deviants, evoked an MMN activity, indexing orientation towards the violation in the acoustic signal. In particular, rising pitch deviants, either on a stressed syllable or at the boundary of a word, engendered the most pronounced MMNs. This shows that pitch rises attract more involuntary attention. Falling pitch deviants elicited less pronounced MMNs, followed by a pronounced P3. This indicates that pitch falls also attract some involuntary attention which, however, ultimately leads to conscious attention orientation. This suggests that rising pitch, as an acoustically salient cue, causes an auditory looming effect at the pre-attentive stage, whereas falling pitch appears to be interpreted as linguistically prominent information within the list context. Thus, its processing is affected by the contextual meaning, which activates conscious attentional mechanisms. The results of this study appear to differ from Hsu et al. (2015) (for a detailed review of this study, see Section 2.4 and Section 3.3) in that Hsu et al. found a P3 response to rising pitch changes at a speaker's normal pitch level (i.e., not elevated).

In what follows, I suggest that the current findings do not, in fact, contradict those of Hsu et al. Crucially, the stimuli used in the current study were different from those used in Hsu et al., as the current study used real words as opposed to a simple /a/ (at normal and resynthesised elevated pitch levels) and sine waves. An additional difference (also described in Section 3.3) is that the rise or fall in the Hsu et al. stimuli involved a change in pitch from one stimulus to the other, i.e., across stimuli. By contrast, in this study, the rise or fall appears within stimuli. Hsu et al. suggest that sudden pitch rises in speech demand more attentional resources than sudden falls, and their presence in speech (in comparison to simple sine waves) activates additional conscious processing mechanisms. In the present study, where the speech signal is the only signal that listeners encounter, I conjecture that it is the linguistic/contextual meaning that draws voluntary attention, rather than speech per se. In other words, one could argue that in both Hsu et al. and in the current study, P3 appears to be elicited by some kind of available “meaning”, and not so much by pitch direction. In Hsu et al., “meaning” emerges from speech in comparison to the meaningless sine waves. It is the sudden rising and not the sudden falling pitch that evokes the P3 because rising pitch is acoustically more salient, but if it was only the direction of pitch and not the additional information of “meaning”, then pitch rises in sine waves would have elicited a P3 as well. In the current study, “meaning” arises from the context. The rising deviant condition, although acoustically salient – and linguistically prominent, as in German prominence is often encoded through intonational rises – happens to be contextually inappropriate, and thus contextually “meaningless”, just like

4 *Unravelling the role of intonation in attention orienting using ERPs*

the sine waves in Hsu et al. By contrast, the falling deviant condition, although acoustically less salient, is contextually appropriate as it transmits the linguistic meaning of the list context. Such a list context may not be as available to speakers of Mandarin, where local f_0 changes are affected by lexical tone.

It has further been shown that different degrees of prosodic prominence trigger signal-driven processes to a different extent (e.g., Röhr et al. 2021). This is also evident in the current results. Specifically, a positive relation between prosodic prominence and MMN activity is observed: as prosodic prominence increases (rises being more prominent than falls), MMN activity intensifies. Nonetheless, the P3 response appears to be unaffected by the prosodic prominence level of the deviant. Instead, it is context-induced. Specifically, the context of the list in the case of the falling deviant (i.e., a sequence of rises) appears to trigger an anticipation of the end of the list, typically marked by a falling contour. Although the falling deviant is acoustically less salient than the rising deviant, it activates conscious attentional mechanisms as reflected in the P3. Thus, in real-time processing of sequentially repeated stimuli, the amount and level of attention allocated to the deviant stimulus appear to be determined by a combination of signal- and context-based properties, when contextual meaning is available. In turn, when contextual meaning is unavailable or inappropriate, signal-inherent properties guide attention orienting. It thus appears that the meaning of the sequence shifts the prominence status of the contours which, in turn, shifts the stage of attention orientation, activating different routes in processing (pre-attentive/conscious, involuntary/voluntary).

Overall, the current results show that the processing of rising and falling pitch contours produces distinct brain responses, which are claimed to be related to two core attentional processes: the automatic involuntary attention switch at the pre-attentive stage (reflected by MMN generation) and the voluntary attention orientation at the conscious stage (reflected in the P3 signature). Pre-attentively, signal-based cues appear to be fundamentally important. At this processing stage, pitch rises, as the most acoustically salient events, attract the most attention. This highlights the pivotal role of pitch rises not only in cognition, as previous MMN studies have shown (e.g., Näätänen et al. 1978, Alain et al. 1994, Doeller et al. 2003), but also in spoken language, suggesting that there is something intrinsic in their acoustic signal, regardless of whether rises take the form of a pure sine wave, a speech sound in isolation or speech in context. This is likely due to the rising acoustic properties being so salient that they are able to warn and prepare the listener's nervous system about an important event happening in the environment, activating basic attentional resources that in turn elicit automated or appropriate adaptive responses (these adaptive responses have been described

by Sokolov 1963b as reflexes; for more see Section 2.1). In spoken communication, this entails orienting a listener's attention towards the most important part of the uttered message, which is crucial for effective interpretation and speech planning in drawing listener attention to an upcoming turn. The essential nature of signal-based cues for the pre-attentive processing stage has also been highlighted by studies reporting that not only salient acoustic contrasts, but also the timing of the acoustic cues are decisive for evoking attention-related brain responses pre-attentively (e.g., Ren et al. 2009, Tsang et al. 2011, Li et al. 2018, Röhr et al. 2021).

Further, the findings of the current study manifest that conscious, voluntary attention is modulated by the meaning that intonation encodes in a given context, and not by pitch direction itself (for similar attentional processes modulated by context-driven cues, see Liu et al. 2016). Here, the conscious processing stage is activated by pitch falls. Although pitch falls are not salient enough and are less prominent in the prosodic prominence hierarchy (e.g., Baumann & Röhr 2015), it is evident that (in this case) language experience and context-driven expectations overwrite the signal-induced properties (see Bishop 2013). In other words, acoustic saliency can be overridden by expectations that emerge from context, making pitch falls highly relevant and thus linguistically prominent in the list context. This finding is in line with studies showing that signal-driven cues can be overwritten at later processing stages by prior linguistic- or discourse-based expectations, as well as by recent speech experience (e.g., Bishop 2013, Kakouros et al. 2018, Roettger & Rimland 2020, Ventura et al. 2020, Bornkessel-Schlesewsky et al. 2022).

4.4.2 The role of the phonological status of the rise (and the fall) in attention orienting

Another question this study has advanced concerns the phonological status of the rises and falls, that is, their position in the prosodic structure and the contribution of this status to the observed attentional processes. Crucially, the results of this study demonstrated differences in the magnitude and latency of the MMN activity when comparing accentual to boundary contours (see Figure 4.11). Specifically, when comparing accentual to boundary rises, accentual rises exhibit the more pronounced MMN. In turn, when comparing accentual to boundary falls, boundary falls show a prolonged MMN latency as opposed to the very short activity evoked by accentual falls. I argue that during online processing, intonation contours, as a complex part of the speech signal, are processed holistically, meaning that attention is not oriented necessarily towards a specific point in the f_0

4 Unravelling the role of intonation in attention orienting using ERPs

trajectory. Rather, it is the pitch contour direction that modulates attention orienting (drawing attention to a higher level of linguistic representation such as a – putative – phrase).

To illustrate this point, consider the rising/falling contours investigated: both accentual and boundary contours were realised on sequentially presented trisyllabic lexical items (like *Banane* ‘banana’). Hence, the domain of the realisation of the pitch contour in this study is the word; thus, each word is an intonational phrase on its own. Therefore, a complete pitch contour on every word consists of both a pitch accent and a boundary tone. Specifically, the pitch configuration which in this study is referred to as *accentual rise* is followed by a high boundary at the end of the word, whereas the pitch configuration which is referred to as *accentual fall* is followed by a low boundary at the end of the word. Likewise, the *rising boundary* is preceded by a low accent, whereas the *falling boundary* is preceded by a high accent. Hence, by considering the entire pitch configuration as the current word stimuli unfold, we can better understand whether a specific part of the contour is relevant for attracting (more) attention or not. The current results indicate that the structural position of the pitch event (accent vs. boundary) has a secondary role in attentional processes. Specifically, it was found that

- i. a rising pitch contour is globally more successful in attracting attention than a falling one, regardless of the position of the rise or the fall, while secondarily,
- ii. a rise on the stressed syllable leads to more attention than a rise on the final unstressed syllable, and
- iii. high pitch on the stressed syllable in falling contours (H^*) attracts more attention than falling pitch on the stressed syllable ($H + L^*$).

The finding that accentual rises induce a greater MMN effect, attracting more attention than boundary rises, is not surprising. It has already been claimed in previous work on prosodic prominence that accentual rising contours (and especially steep rising pitch accents) are the most prominent contours in the prosodic prominence hierarchy and that they demand more attentional resources than falling pitch accents (e.g., Baumann & Röhr 2015, Röhr et al. 2021). Let us now zoom in on the finding that accentual rises attract more attention than boundary rises, considering it in light of the entire contour and the periodic energy that characterises these rising contours (for results on periodic energy, see Section 4.2.2). In the accentual rising contours, pitch starts low, already rises quite steeply during the stressed syllable, and remains high towards the last syllable,

and thus at the end of the word (an appropriate analysis following the German Tones and Break Indices (GToBI) annotation scheme, would be: L + H* H-%; see Grice et al. (2005)). The rising pitch movement takes place on the lexically stressed syllable, which has high periodic energy, making the pitch strongly transmitted on that syllable. See Figure 4.1 for mean and individual intonation contours, Figure 4.3 and Figure 4.4 for measures related to periodic energy. Now, in boundary rising contours, pitch starts mid-level, remains at this level or falls slightly during the stressed syllable, and rises towards the end of the word. In GToBI, this would be annotated as L* L-H%. The rising part of the contour is restricted to the final unstressed syllable, which has considerably lower periodic energy than the stressed syllable. Thus, the f0 in the boundary rise condition is transmitted more weakly than in the accentual rise condition. Therefore, although both rising contours attract more attention than the falling ones, within the rising category, the accentual rising contours are produced with more periodic energy and thus attract more attention than the boundary rising contours.

Turning to the comparison of accentual and boundary falling conditions, a more sustained MMN effect is evoked by boundary falls rather than accentual falls. Consider that, in the realisation of the accentual falling contours, pitch starts high at the beginning of the word (enabling it to fall) and falls throughout the stressed syllable. It then continues on the same low level until the end of the word (in GToBI, H + L* L-%). In the realisation of the boundary falling contours, pitch starts on a relatively mid-level and rises slightly towards the stressed syllable, enabling it to fall at the end of the word (in GToBI, H* L-%). The high accent on the stressed syllable (and before the fall at the boundary) leads to higher periodic energy on this part of the signal in comparison to the accentual falling contour, where pitch is already falling during the stressed syllable (this is conversant with the finding that H* is more prominent than H + L*; see Baumann & Röhr 2015). Therefore, the high pitch accent and the amount of periodic energy in the signal before the boundary fall potentially contribute to the perception of the boundary falling contour condition as more prominent, leading to greater attention as compared with the accentual fall.

One could argue that this constitutes evidence for the attention orienting function and the prominence value of the accent, and thus the structural importance of the stressed syllable. However, remember that in comparing boundary rising to boundary falling contours, boundary rises were found to attract more attention than falls, although the stressed syllable in the former contour bears a low pitch accent, as opposed to the boundary fall that is preceded by a high pitch accent. It appears, therefore, that what takes place (in terms of f0 movement) at discrete prosodic positions (head/edge) is not sufficient to orient attention on its

4 Unravelling the role of intonation in attention orienting using ERPs

own. In contrast, it is the entire contour that guides attention. Prosodic positions appear to have a supplementary/secondary role in the modulation of attentional resources. This becomes evident when the investigated pitch contours are arranged according to their elicited MMN effects, assuming a decrease in MMN effect (and thus attention attraction) from left to right:

$$\begin{array}{c} \text{RISE} [\text{accentual rises (L + H* H-)} > \text{boundary rises (L* L-H)}] \\ > \\ \text{FALL} [\text{boundary falls (H* L-)} > \text{accentual falls (H + L* L-)}] \end{array}$$

First and foremost, a rising pitch configuration globally attracts more attention than a falling one. Second, and within rising categories, when the rise occurs in different structural positions, it attracts more attention when it coincides with the head/stressed syllable (in the accentual rising condition) compared with when it occurs at the boundary (in the boundary rising condition). Now, when the rise occurs at the same structural position, it attracts more attention if it is a steep rise (L + H* in the accentual rising condition) as opposed to a shallow rise or just (level) high pitch (H* in the boundary falling condition). Finally, within falling categories, high pitch preceding a fall on the accented syllable (in the boundary falling condition) attracts more attention than a simple fall (in the accentual falling condition). These results are in line with what has previously been reported on prominence marking in German (e.g., Baumann & Röhr 2015, Baumann & Winter 2018). Crucially, it was shown here that these subtle prosodic differences are reflected in the pre-attentive MMN response.

4.4.3 Some implications for the neural architecture of language

The neural architecture of language perception is complex and dynamic (for discussions, see among others, Gandour et al. 2004, Hickok & Poeppel 2007, Assaneo et al. 2019). It involves two fundamentally different neural mechanisms, a signal-based mechanism and a meaning-based one, expressed at distinct processing stages. On the one hand, the current results show that the intrinsic properties of the sensory input, that is, signal-driven cues (Assaneo et al. 2019 refers to this as the *intrinsic auditory mechanism*), are essential for speech perception at the early pre-attentive processing stage, with rises taking priority over falls precisely because of their acoustic saliency. The fundamental role of signal properties, and thus the special role of rises at the pre-attentive processing stage (feeding the signal-based mechanism) is also shown in previous MMN studies to unexpected

sound changes (for a review of studies, see Chapter 2, and among others, Näätänen et al. 2019) but also in studies investigating the neural processing of linguistically meaningful variations, both at the lexical and the postlexical level (e.g., Ren et al. 2009, Tsang et al. 2011, Li & Chen 2018, Röhr et al. 2021). On the other hand, the current findings suggest that at a later, conscious, processing stage, the linguistic functions of the stimuli (called *top-down/externally driven mechanism* by Assaneo et al. 2019) modulate speech perception (feeding the meaning-based mechanism). Specifically, the construction of meaning was found to attract voluntary attention towards meaningful aspects, here reflected by the use of words and the minimal context of list intonation. This is in line with findings from previous studies showing that top-down activities are decisive for the activation of later processing stages (see, among others, Ren et al. 2009, Liu et al. 2016, Assaneo et al. 2019, Röhr et al. 2021).

Overall, it appears that the neural architecture of the auditory cognition and spoken language share a common pool of mechanisms that are crucial for the orienting response in both auditory and speech processing. Chapter 6 brings together the results from this and the following chapter in discussing implications for the neural architecture of general cognition and spoken language.

4.5 Summary

The present study is a novel attempt to unravel the neural mechanisms that underlie attention orienting towards unexpected linguistic intonational changes by revisiting the idea of an attentional bias towards pitch rises (as opposed to pitch falls) and extending it from a general cognitive level (auditory looming) to a linguistic one. This study shows that, in a linguistic context, the amalgamation of different cues evokes qualitatively and quantitatively distinct neural responses tied to two core attentional mechanisms:

1. the involuntary attention switch, a mechanism at the pre-attentive processing stage which is reflected in MMN elicitation, and
2. the voluntary attention orienting, a mechanism at the conscious processing stage which is reflected in the P3 signature.

In its most concise form, the main finding of this study is that, in spoken language, rising intonation takes on a special role in attracting involuntary attention, whereas contextual meaning is essential for voluntary attention orienting.

4 Unravelling the role of intonation in attention orienting using ERPs

Rising pitch evokes the largest MMN, indicating that it leads to a greater involuntary and automated attention switch compared with falling pitch. This holds regardless of the phonological association of the rise in the prosodic structure (head vs. edge). It thus appears that there is a biological basis for the cross-linguistic use of rises for attracting attention towards informative parts of the message – even though rises are grammaticalised differently across languages in being encoded as either pitch accents or edge tones. Furthermore, the appropriateness of the intonational pattern in a given context and, specifically in this study, in the list context, is decisive for voluntary attention orienting. Here, falling pitch, although acoustically less salient than rising pitch, engenders an additional P3, indicating that the contextual meaning prevails over or even “cancels out” the signal properties (for a discussion on the interplay between sensory input and top-down activities in language processing, see Bornkessel-Schlesewsky et al. 2022).

Overall, the role of rises is fundamentally important, not only for auditory cognition but also for (successful) language communication, because of their intrinsic acoustic properties that activate involuntary attentional resources, pre-attentively, regardless of whether they are carried by a simple sine wave, a speech sound in isolation, or meaningful speech. However, in spoken language, the acoustic signal is not the only source of information; hence, the importance of rises is mitigated by contextual meaning, which appears to activate voluntary attention as it is required for modulating/updating conscious processing stages and mental representations during language comprehension. Capitalising on the current findings, the next chapter delves further into the relevance of domain-final (the reflex of phrase-final edge tones) intonational rises (and falls) for attention orienting in spoken language by measuring pupil dilation response (PDR), a rigorous psychophysiological index of auditory attention orienting (see Chapter 2).

5 Indexing the attention orienting function of domain-final rises using pupillometry

A fundamental characteristic of the human cognitive system is its ability to withdraw attention from a currently important activity and orient it towards a source of information that may be currently irrelevant, but valuable enough for further assessment. For example, an unexpected, rare, or new sound that deviates in some property from the current auditory environment may capture attention, prompting an involuntary switch of the attentional resources towards this deviant auditory event. In Chapter 4, I demonstrated the special role of rising intonation in attention orienting towards unexpected changes in spoken language in an event-related potential (ERP) study. More specifically, I showed that the attention orienting function of rising intonation in speech holds regardless of whether the rise was accentual or associated with an edge tone. Capitalising on these findings, the current study delves further into the relevance of domain-final¹ intonational rises and falls (which are the reflex of phrase final edge tones) for attention orienting in German. Using a *changing-state* oddball paradigm, in which auditory sequences of sequentially ordered (i.e., seriatim) ascending numbers (*standards*) are occasionally interspersed with an out-of-the-sequence number (*deviant*), this study investigates whether domain-final rising intonation takes on a special role in attention orienting by measuring listeners' pupil dilation response (PDR).²

5.1 Motivation

Over the years, many studies have been concerned with the mechanism underpinning attention orienting. This work has paved the way for the development of many accounts. As reviewed in Chapter 2, one of the mechanisms that has been

¹Hereafter, I use the terms *domain-final*, *edge*, and *boundary tone* interchangeably.

²A part of the data reported on this chapter has been presented at the 46th Annual Meeting of the Cognitive Science Society (CogSci 2024) and published in the conference proceedings (Lialiou, Harris, et al. 2024).

suggested as underlying auditory attention orienting is *expectancy violation*: the auditory system is able to develop expectancies by detecting regularities in the sound environment, and thus predict upcoming sound events. When a deviant sound occurs instead of an anticipated event, it attracts attention, initiating an orienting response (OR; for more on the mechanisms of attention orienting, see Section 2.1 and Section 2.3). It has been shown that this OR finds expression in multiple physiological indices, such as heart and respiratory rates, and electrodermal, vasoconstrictive, neural, as well as pupillary reactions (for more, see Section 2.5).

As discussed in Chapter 2, pupillary responses, and more specifically, PDR has mainly been used as a correlate of cognitive processes like listening effort, heightened attention, expectancy violation, and memory consolidation (for review, see Winn et al. 2018). PDRs have also been used in linguistic studies assessing cognitive resource allocation towards mismatches between prosody and focus in discourse processing (e.g., Zellin et al. 2011); attention and linguistic processing difficulty towards grammatical, syntactic, and semantic mismatches (e.g., Demberg 2013, Demberg & Sayeed 2016); affective arousal towards expectancy violations such as rhyme violations in poetry processing (e.g., Scheepers et al. 2013); attention and cognitive load towards mismatches between prosodic and syntactic grouping (e.g., Jesse & Jun 2019); and attention allocation towards unpredictable sentence continuation and age-effects (e.g., Häuser et al. 2019), among others. More importantly, PDR has also been identified, by a growing number of studies, as a rigorous (psycho)physiological index of both involuntary and voluntary auditory attention orienting, highlighting a striking resemblance between the PDR composites and the neural MMN/P3 responses observed using ERP measures (for more on PDR as a proxy for attention orienting as well as on the link between neural and pupillary activity, see Section 2.5). Specifically, as discussed in Chapter 2, pupillometry provides a high resolution of temporal information, making it possible to measure cognitive activity related to attentional processes over time. The latency of attention-related PDRs has been reported to be indicative of the processing stage in which the attentional mechanism is activated: transient dilations reflect pre-attentive processes, while more prolonged responses indicate processes at the conscious level (see Strauch et al. 2022).

Pupillometric studies on auditory attention, using oddball paradigms, have further reported that PDR is not only affected by the presence of an auditory deviant, but is also sensitive to the intensity of the physical characteristics of that deviant sound. In other words, the more salient the acoustic properties of a deviant sound, the larger the magnitude of the change in the PDR (e.g., Liao

et al. 2016, Wetzel et al. 2016, Marois et al. 2018). These findings align with neurophysiological and linguistic research on auditory and speech perception. In auditory perception, the saliency of an event is essential in attracting attention: the greater the rise in amplitude or pitch on a deviant sound, the greater the orienting response. Likewise, in speech perception, intonational rises are used for attracting interlocutors' attention when asking questions, directing listeners' attention towards the most important part in an utterance, and even guiding attention in serial recall tasks (for a review of auditory cognition, see Section 2.4; for a review of language perception, see Section 3.3 and Section 3.4).

The current study examines how attention orienting in response to deviances in numeric sequences is modulated by the intonation pattern that these deviances feature. Listeners' pupil size, and specifically, PDR, is used as a proxy for attention orienting. As discussed in Chapter 3, intonational events are phonologically anchored to specific positions in the prosodic structure, forming either pitch accents (anchored to stressed syllables), or edge tones (anchored to edges of constituents). In the autosegmental-metrical theory of intonational phonology (e.g., Ladd 2008), edge tones have been strictly associated with a phrasing function, while pitch accents have been associated with a prominence-cueing function. Serial recall studies (e.g., Savino et al. 2020, Röhr, Savino, et al. 2022, Grice et al. 2024) have provided some evidence against this strict dichotomy in reporting that rising edge tones attract attention by lending prominence to the whole domain they delimit, as reflected in listener recall performance. Building on these serial recall studies as well as the finding that rises associated with the edges of constituents induce an attention orienting response in a similar way to accental rises (reported in Chapter 4), the present study focuses on the processing of domain-final rising and falling intonation in German (RQ1 [rise]), and further investigates the attention orienting function of rises at the edges of constituents (RQ2 [edges]).

Moreover, considering the relevance of meaning for attention orienting in spoken language reported in Chapter 4, the present study maintains and refines the minimal context of lists employed in the previous chapter, this time using auditory numeric sequences consisting both of seriatim ascending numbers (*standards*; e.g., 21 22 23 24 25 26 27 28...) and occasional out-of-the-sequence numbers (*deviants*; e.g., 25 in 21 22 23 25 26 27 28...). Numeric sequences were selected for this study as they inherently give rise to the linguistic context of a list, allowing for expectancy formations as well as expectancy violations, which are both related to the forthcoming numbers. This provides an optimal context for studying attention orienting. Moving to the language-specificity of this context (RQ3 [expectations]), standard numbers feature shallow falling intonation (hereafter,

neutral intonation), a pattern that can also be found on non-final items of a sequence or list in German, while deviant numbers are produced with one of the following three intonational patterns: neutral intonation, domain-final rises, or domain-final falls. Domain-final rises and falls are functionally distinct: rises indicate continuity; falls denote finality (e.g., Grabe 1998, Baumann & Trouvain 2001, Chen 2003, Peters 2018). However, both types of edge tone can mark the end of either small or large units in a sequence, and they thus fulfil a similar function (for more on German list intonation, see Section 3.5). Therefore, the naturalistic pitch contours used in this study for the realisation of both standard and deviant numbers simulate a natural linguistic context.

Assuming that attention orienting is indexed, at least in part, by PDR excursions, the prediction is that the presentation of a deviant number (as 27 in 23 24 25 **27** 28...) will disrupt the anticipated pattern. This prediction is based on the claim that the attention orienting response is underpinned by an expectancy violation mechanism (for more, see Chapter 2). The disruption of the anticipated pattern caused by the deviant number will thus capture attention, which in turn will induce an excursion in PDR. Based on the enhanced orienting function of rising pitch compared to falling pitch (for reviews, see Section 2.4, Section 3.3 and Section 3.4, and for current results, see Chapter 4), it is further predicted that a domain-final rise in deviants will result in greater disruption, attracting more attention and thus inducing a more robust PDR compared to deviants produced with domain-final falls or with neutral intonation.

Now, individuals may vary in the use or control of attentional resources towards incoming information. These individual differences may arise from the use and activation of different cognitive functions interacting with the attentional system. Studies endeavouring to better understand individual variability during cognitive or language processing usually examine cognitive functions like working memory capacity (WMC), processing speed, and executive processes (e.g., inhibition, shifting, updating), among others (for review, see Frischkorn et al. 2022). At a more exploratory level, this study attempts to unravel whether listeners' individual cognitive profiles affect the processing of the introduced arithmetic deviances produced with different intonational patterns, as a way of better understanding the attention orienting function towards them. In this endeavour, selective attention, in terms of inhibitory ability, processing speed, and WMC are selected as proxies for measuring individual cognitive variability. Inhibition may be one of the most crucial cognitive operations for better understanding the mechanism of attention orienting towards deviant events across individuals. Inhibition reflects a listener's ability to suppress irrelevant or unimportant information from breaking into the current attentional focus. Thus, whereas

individuals with better inhibitory ability might be better at suppressing unimportant auditory deviances, thus saving attentional resources, individuals with lower inhibitory capacity might be more susceptible to switching attention towards auditory deviances. Further, there is a consensus in the literature on individual variability that processing speed (although a diverse term; for discussion see Frischkorn et al. 2022) is essential in explaining varying processing patterns among individuals. Therefore, processing speed might also contribute to a better comprehension of the orienting response by exploring whether processing speed aids or impedes the evaluation of deviants, for example, on the basis of signal-driven cues like different intonation patterns (as in this study). Lastly, WMC is one of the most frequently investigated cognitive measures of individual variability. While WMC has been found to be a strong factor in individual variability in cognition, its contribution to attentional control is less clear (for discussion, see e.g., Keye et al. 2009, Sörqvist et al. 2013). However, some studies have reported a link between working memory (WM) load and mitigation of distractions at the later stages of orienting, such that the greater the WM load, the less susceptible a person is to distractions (e.g., SanMiguel et al. 2008). Thus, in the context of the current study, WMC measures might shed more light on the link between attention-related mechanisms and WMC resistance operations.

Lastly, this study puts forward another exploratory question concerning the contribution of sequence length (if any) in attention orienting. This question was born out of the design of the items, whereby sequence length served the purpose of making the deviant position less predictable throughout the experiment (see Section 5.2.2). Given the exploratory nature of this question, no hypothesis is formed. The present study thus has two aims, a confirmatory one and an exploratory one. The main aim is to test the attention orienting function of domain-final rising intonation, using PDR as a rigorous proxy for attention orienting. In addition, this study puts forward two exploratory questions, one related to the effect of sequence length in attention orienting, and another exploring individual-specific cognitive variability in the activation of attentional resources towards deviances featuring different intonational patterns.³

³Prior to pupil data collection, the design of the stimuli was tested in a pilot study, ensuring that the introduced deviances were indeed perceptible. This pilot study is reported in Appendix A.

5.2 Methods

5.2.1 Participants

Sixty native speakers of German (54 female, 6 male), aged between 19 and 38 years (mean age = 22.6 years, $SD = 3.3$) with normal or corrected-to-normal vision participated in this study. Participants provided written informed consent in accordance with the Declaration of Helsinki and in compliance with the ethics clearance from the Ethics Board of the *Deutsche Gesellschaft für Sprachwissenschaft* (DGfS). Participants received reimbursement for their participation (either course credit or monetary compensation). None of them reported any speech, hearing, or neurological impairments.

5.2.2 Speech materials

The auditory stimuli comprise sequentially ascending ordered lists of numbers, consisting of a set of 17 numbers (medium-length sequence) or 22 numbers (long sequence). They were combined with three different prosodic realisations (neutral, rise, fall) on the deviant number. In total, 36 unique experimental numeric sequences were constructed for this study, with 6 different numeric sequences per prosodic condition and sequence length. The experimental items were combined with 36 unique filler items, which did not include a deviant number. Figure 5.1 illustrates instantiations of filler and experimental materials.

The experimental items introduced arithmetic deviances, i.e., a number out of sequence, called deviant. To achieve this, one or two consecutive numbers were omitted from the sequence. Specifically, 18 out of the 36 numeric sequences (i.e., 3 out of the 6 numeric sequences, per prosodic condition and sequence length) introduced the deletion of 1 consecutive number, and the remaining 18 numeric sequences introduced the deletion of 2 consecutive numbers. The controlled variation in deletion of one or two consecutive numbers served to increase the difficulty of the task. To control for potential effects of deviant position in the sequences, the deviant was introduced at two different positions: position 11 in the medium sequence length and position 16 in the long sequence length, as shown in the second panel of Figure 5.1. Sequence length served the purpose of making the position of the deviant less predictable throughout the experiment.

The rise and fall prosodic conditions involved domain-final pitch movements, reflecting phrase final High and Low edge tones, respectively. The prosodically neutral condition served as a baseline. The experimental sequences included numbers between 22 and 99 which consisted of either two (e.g., 50 *fünfzig* [fʏnftsiç]), four (e.g., 52 *zweiundfünfzig* [tʃvaɪʏntfʏnftsiç]) or five syllables (e.g., 57

Filler items																						
# position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17					
sequence	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96					
# position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
sequence	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
Experimental items																						
<i>deviant</i>																						
# position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17					
sequence	37	38	39	40	41	42	43	44	45	46	48	49	50	51	52	53	54					
<i>deviant</i>																						
# position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
sequence	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	61	62	63	64	65	66	67

Figure 5.1: Example filler and experimental materials illustrating the different lengths of the numeric sequences used, as well as the two different deviant positions in the experimental items.

siebenundfünfzig ['zi:bənuntfʏnfstɪç]), always with primary stress on the first syllable, allowing enough time for the different intonation contours to unfold. For the deviant numbers, 32 of 36 consisted of four syllables, and the 4 remaining numbers consisted of five syllables.

An example of a numeric sequence for each of the three prosodic conditions is depicted in Figure 5.2. In the rising condition, all standards in the sequence were produced with the same intonation, that is, with neutral intonation, typically used on non-final items of a sequence or list in German. The deviant was realised with a boundary rising intonational contour. In the falling condition, similarly to the rising one, all standards in the sequence were produced with the same neutral intonation, and the deviant was realised with a boundary falling intonational contour. A neutral condition, in which both standards and deviants in the sequence were produced with the shallow falling intonation, served as the baseline condition. While both boundary rises and falls can mark the end of small or large units in a sequence, they are functionally distinct: rises indicate continuity; falls indicate finality. Across all three conditions, the last number of the entire sequence of each trial was realised with a boundary falling intonational contour in order to mark the end of that trial.

5 Indexing attention orienting of domain-final rises using pupillometry

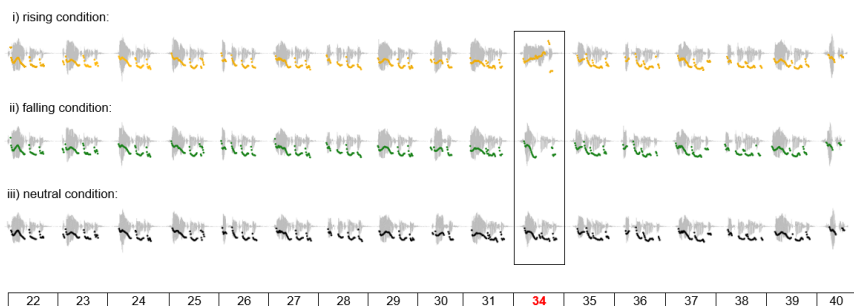


Figure 5.2: Speech waveform & f0 contour of a sample experimental trial sequence per condition.

The filler items were constructed without a deviant in the sequence, enhancing the creation of the memory trace formation of the sequentially ascending ordered numbers. The filler items consisted of a different range of numbers compared to the experimental ones to ensure variability in sequence construction. Numbers ranged between 2 and 99, and consisted of either one, two, four, or five syllables, always with primary stress on the first syllable. Filler items were comparable to the experimental items with regard to the prosodic conditions in order to ensure that participants could not identify the deviant in the experimental items just by tuning into the prosodic marking of deviants. Out of the filler items, 12 highlighted a number in the sequence with a boundary rising intonational contour (comparable to the rising condition), another 12 items highlighted a number in the sequence with a boundary falling intonational contour (comparable to the falling condition), and a further 12 items were comparable to the neutral condition. Fillers differed from experimental items in that the position of the prosodically distinct number in the sequence was fully randomised, in order to ensure that participants would not be able to identify a particular position in the sequence which differed prosodically from the rest.

Participants were presented with all 72 items (36 experimental and 36 fillers; 12 for each of the three prosodic conditions) in a counterbalanced design. Specifically, items were distributed in three lists. Each list contained all items and conditions, but never the same item across conditions. The 72 items were further distributed across three blocks with 24 items each (12 experimental and 12 fillers). The order of the items in each block was pseudo-randomised, with at most three consecutive experimental items but never from the same condition. To control for systematic order and frequency effects potentially induced by the exposure

to block and/or item order, the fully counterbalanced lists were created with different block order and item, so that each list presented all items and blocks but never the same item across blocks and never the same block order. Each participant heard only one of the lists (the exact distribution of the lists as well as all items have been made available on the Open Science Framework (OSF) platform [<https://osf.io/nhy4c>]). All stimuli were produced by a phonetically trained 38-year-old female native speaker of German and recorded with a sampling rate of 44.100 Hz and 16-bit resolution (mono). To ensure natural speech production of the items, first, the speaker produced all numbers from 0 to 100 in separate blocks as a function of prosodic condition (e.g., neutral prosody: 0, 1, 2, 3,...100; rising prosody: 0, 1, 2, 3,...100; falling prosody: 0, 1, 2, 3,...100). Subsequently, all number renditions were cut from each block, saved as individual audio files, and concatenated into the different numeric sequences using Praat (Boersma & Weenink 2024). The inter-stimulus interval between spliced numbers was 100 ms. The average duration of the medium-length and long sequences was 24.86 sec and 25.21 sec, respectively. All stimuli used in the experiment were normalised at -23 LUFS but not manipulated further.

5.2.3 Acoustic characterisation of speech material

For the acoustic characterisation of the deviant numbers in the sequences, the relative Delta f_0 (Δf_0) metric from the *ProPer* toolbox was employed (Albert et al. 2018, Cangemi et al. 2019, Albert et al. 2020, Albert 2023), as described in Chapter 4. Prior to the *ProPer* analysis, f_0 contours were extracted and corrected manually in Praat (Boersma & Weenink 2024), using a customised version of *mausmooth* (Cangemi 2015). The *ProPer* analysis was conducted on the basis of syllabic units. Scripts and data tables of the current analysis have been made available on OSF (<https://osf.io/nhy4c>). The measure of Δf_0 traces the f_0 trajectory across syllables, using both f_0 and periodic energy, indicating f_0 changes from syllable to syllable by calculating the difference from the previous one. For more details about the metric please refer to Section 4.2.2. Table 5.1 presents means and standard deviations for Δf_0 per syllable across target numbers for each prosodic condition, as well as the total duration of the deviant numbers.

Figure 5.3 depicts relative Δf_0 values per syllable as a function of prosodic condition across quadrisyllabic deviant numbers (the pattern is the same in pentasyllabic deviant numbers, as can be seen in Table 5.1). In the rising condition (see depiction in yellow), mean Δf_0 starts at a mid-level and rises shallowly from the first to the second syllable, then remains on the same level until syllable three, and steeply rises towards the last syllable, i.e., the right edge of the word.

5 Indexing attention orienting of domain-final rises using pupillometry

Table 5.1: Mean and standard deviation values (in brackets) for relative Δf_0 values per syllable across items for each condition as well as for the total duration of items.

measurement		rising	falling	neutral
relative Δf_0	<i>quadrisyllabic</i>			
	1	0.97 (4.74)	5.59 (3.92)	2.82 (8.90)
	2	9.33 (3.36)	-17.9 (7.79)	-8.23 (4.93)
	3	7.04 (3.18)	-28.2 (12.4)	-11.5 (11.3)
	4	21.7 (7.12)	-15.8 (7.34)	-14.2 (15.0)
relative Δf_0	<i>pentasyllabic</i>			
	1	0.88 (1.38)	3.88 (1.89)	1.47 (2.07)
	2	3.00 (0.79)	-4.33 (0.87)	-1.16 (2.09)
	3	4.75 (3.13)	-14.7 (2.99)	-8.23 (1.19)
	4	5.24 (0.65)	-23.7 (7.41)	-4.13 (1.67)
	5	27.2 (13.3)	-15.9 (7.32)	-10.8 (13.8)
duration [ms]	item	1126 (59.64)	1127 (86.06)	1098 (61.64)

Figure 5.4 presents instantiations of the deviant f_0 contours per prosodic condition, analysed in the autosegmental-metrical model of intonation laid out in the German Tones and Breaks Indices annotation scheme (GToBI; Grice et al. 2005). In GToBI, the aforementioned rising contour is $H^* \wedge H\%$. In both falling and neutral conditions (see Figure 5.3, depiction in green and black, respectively), mean Δf_0 starts somewhat higher than in the rising condition, and gradually falls from the first to the last syllable. The difference between the falling and neutral condition is that in the falling condition, the first syllable is slightly higher, and the drop from the first to the second, and from the second to the third syllable, is steeper than in the neutral condition. The drop towards the last syllable is smaller in the falling than in the neutral condition. In GToBI, the neutral and the falling conditions are labelled as $H^* L-$ and $H^* L\%$, respectively (see Figure 5.4, depiction in green and black, respectively). The larger boundary in the falling condition explains the gradient difference in the meaning between the two contours. Specifically, the neutral contour, because it does not fall as steeply as the falling contour, marks the end of a unit less clearly, and thus can be featured on non-final items in a sequence. The falling contour, in contrast, unambiguously marks the end of a unit indicating finality due to its steep fall and the extra-low

pitch towards the end. Lastly, the rising contour can mark the end of a unit as well, but it is functionally different from the falling contour in that it indicates that more is expected (in a following unit).

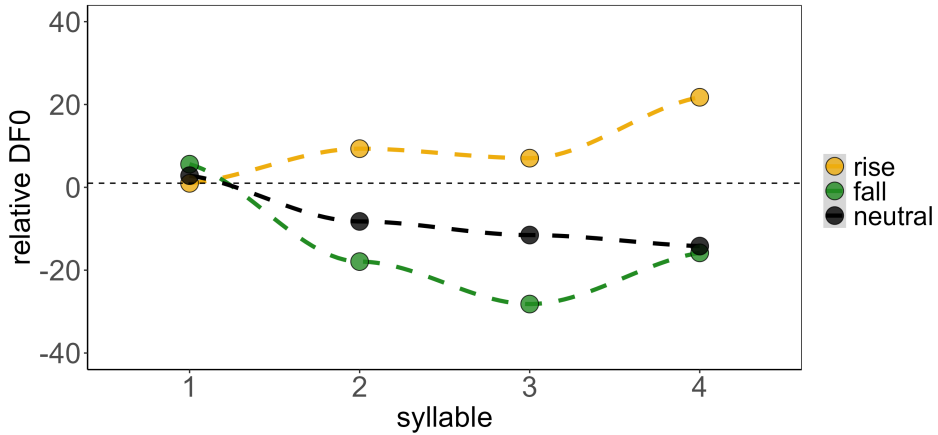


Figure 5.3: Mean relative Δf_0 values per syllable (x-axis) across prosodic conditions in quadrisyllabic deviant numbers.

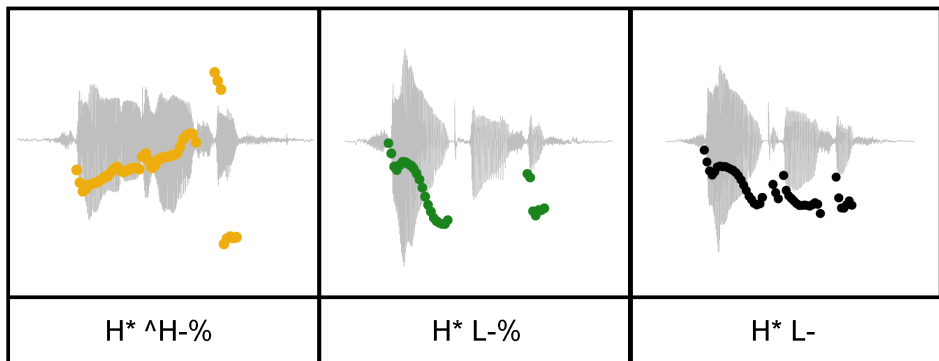


Figure 5.4: Instantiations of deviant f_0 contours per prosodic condition in GToBI.

5.2.4 Experimental procedures

The experimental sessions of the main study followed the exact same structure as the pilot study (see Section A.3). That is, the pupillometric task was performed

5 Indexing attention orienting of domain-final rises using pupillometry

always at the beginning of the session, followed by a battery of cognitive tests, always in the following order: a version of the flanker task (measuring inhibitory ability), a version of the odd-man-out task (measuring processing speed), and a version of the digit span task (measuring WMC; see below for more details on these tasks).⁴

5.2.4.1 Pupillometry

For the pupillometric task, the auditory stimuli were presented using the SR Researcher Experiment Builder (v. 4.595) via loudspeakers, with pupil data time-locked to the onset of each sequence. Pupil size was sampled at 1000 Hz, using an EyeLink 1000 eye-tracker (SR Research Ltd.). Prior to the beginning of the task, the system was calibrated to the dominant eye of each participant, using a 9-point calibration procedure. For all participants, the average luminance measured at the dominant eye was 50 lx.

Participants were seated in the eye-tracking lab in front of a computer monitor and a keyboard. Participants were informed that they would be presented auditorily with numeric sequences but they were naive to the deviances included. In order to keep them engaged with a task, participants had to answer a comprehension yes/no question related to the numeric sequence they heard in 35% of the trials,⁵ by pressing a button indicated on the keyboard. Written instructions were also provided. The experiment started with a practice phase of 5 items consisting of seriatim numbers. Two of them were followed by a comprehension question for which participants received immediate feedback on the screen. The experiment consisted of three blocks of 24 items each. Participants could take an optional short break between the blocks. The experiment lasted approximately one hour.

Every trial started with a drift correction. With the onset of the auditory sequence, a black fixation cross appeared on a grey background and remained on the screen during the whole trial. Participants were instructed to fixate on the

⁴The button press, the flanker, and the odd-man-out tasks were implemented in *OpenSesame* (Mathôt et al. 2012). The digit span task was implemented in *SoSci Survey* (Leiner 2024). In all these tasks, participants could take an optional short break between the blocks, and during the practice phase they received immediate feedback on the screen.

⁵For example, one of the questions was *War die Aufzählung in Zehnerschritten?* “Was the enumeration in steps of ten?” The full set of the questions is provided on OSF. Participants’ mean response accuracy to these questions was 95% for questions related to the experimental items, and 100% for questions related to the fillers (across individuals, response accuracy ranged between 86–100%), indicating high engagement.

black cross on the screen. Twenty-five of the trials were followed by a comprehension question presented onscreen, and participants had to press one of two buttons on the keyboard to indicate their yes/no response. Following the offset of each trial, a grey screen with a black dot was displayed, providing enough time for the pupil dilation to subside. Specifically, participants were instructed that during this screen, they could take time to rest their eyes. Once they were ready to continue, they had to press SPACE to start the next trial. Figure 5.5 depicts a schematic illustration of an experimental trial.

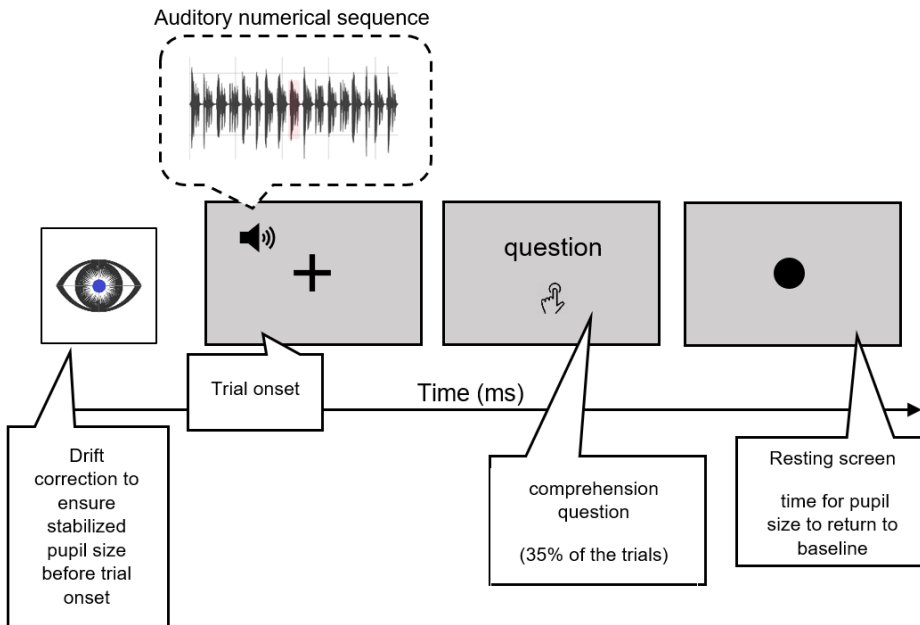


Figure 5.5: Schematic illustration of an experimental trial in the pupil-lometric task.

5.2.4.2 Cognitive test battery

Various cognitive functions may give rise to individual differences in language processing. In the literature, a large number of tasks has been employed to investigate individual differences in the performance of cognitive and linguistic tasks. The cognitive battery employed in this study consists of modified versions of the flanker (Eriksen & Eriksen 1974), odd-man-out (Frearson & Eysenck 1986), and digit span (Wechsler 1987) tasks, endeavouring to measure selective attention in terms of inhibitory ability, processing speed, and WMC, respectively. The

current cognitive battery has been made freely available online for use in other studies (repository link: <https://osf.io/muh9t/>).

5.2.4.2.1 Flanker task

To measure inhibitory ability, a modified version of the flanker task was employed, using arrow stimuli (Eriksen & Eriksen 1974, Anwyl-Irvine et al. 2020). Participants were informed that a sequence of five arrows would appear horizontally in the middle of the screen. Participants were asked to indicate whether the middle arrow in this sequence pointed to the left or to the right by pressing the “S” or “K” button on the keyboard accordingly. They were instructed to place their fingers on the above-mentioned buttons to reduce cognitive load. The task consisted of two conditions: a congruent and an incongruent one. As depicted in the first two rows of Figure 5.6, in the congruent condition, all arrows pointed in the same direction as the middle arrow. In the incongruent condition, the middle arrow pointed in the opposite direction compared to the other arrows in the sequence (see last two rows). Participants were asked to be as fast and as accurate as possible in their responses.

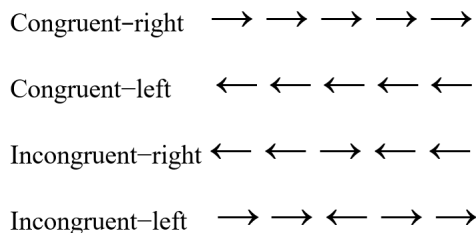


Figure 5.6: Item illustration as a function of condition in the flanker task.

In total, 96 experimental items were constructed for this task (48 congruent and 48 incongruent). Half of the congruent items (i.e., 24 items) presented all arrows pointing towards the right direction (see first row in Figure 5.6), while the other half presented all arrows pointing to the left (see second row in Figure 5.6). Half of the incongruent items (i.e., 24 items) displayed all arrows pointing to the right direction, while the middle arrow pointed to the left (see last row in Figure 5.6). The remaining 24 incongruent items exhibited the reverse pattern, i.e., all arrows were pointing to the left, while the middle arrow pointed to the right (see third row in Figure 5.6). The task started with a practice phase of 12 items. The main experimental phase consisted of four blocks with 24 items each

(25% congruent-right direction, 25% congruent-left direction, 25% incongruent-right direction, 25% incongruent-left direction). Item and block order were fully randomised. Figure 5.7 depicts a trial structure: for each trial, a fixation cross was presented on the screen for 1700 ms. The fixation cross was followed by the presentation of the arrow stimuli. The stimuli remained on the screen until participants pressed one of the two valid buttons (“S” or “K”). Each trial ended with a blank screen for 400 ms. The task lasted approximately 5 minutes.

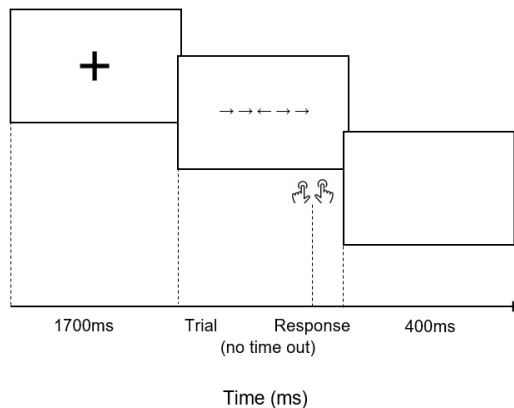


Figure 5.7: Schematic illustration of an experimental trial in the flanker task.

5.2.4.2.2 Odd-man-out task

To measure processing speed, a modified version of the odd-man-out task was employed (Frearson & Eysenck 1986). In the original task, participants were presented with a disposition of three stimuli and were asked to decide whether the first or the third stimulus is more distant in relation to the second one, i.e., which is the “odd-one-out” in terms of proximity. Here, following Diascro & Brody (1994), the original task was adapted by increasing the complexity of the paradigm. Specifically, two levels of difficulty were manipulated: the distance of stimuli disposition (hereafter, spatial condition) as well as the shape (hereafter, gap condition) of the “odd-one-out” stimulus. The stimuli consisted of three hexagons arranged in two odd-man-out conditions. In the spatial condition, the stimuli were arranged in a line, with the odd-man-out stimulus at a spatial distance from the other two hexagons. As depicted in the three top panels of Figure 5.8, three

5 Indexing attention orienting of domain-final rises using pupillometry

spatial “odd-one-out” configurations were possible in this condition. In the gap condition, stimuli were arranged at equal distance from each other in a line, but they differed in that the “odd-one-out” hexagon had a gap on its top side. Again there were three possible locations for the missing side (the right, left, or middle hexagon). The three bottom panels of Figure 5.8 display the possible “odd-one-out” configurations in the gap condition.

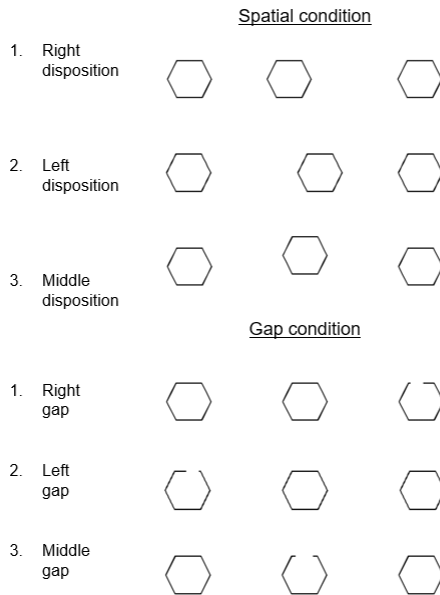


Figure 5.8: Item illustration as a function of condition in the odd-man-out task. The three top panels show the possible locations of the “odd-one-out” stimulus’s disposition in the spatial condition; the three bottom panels illustrate the possible locations of the missing side of the “odd-one-out” stimulus in the gap condition.

Participants were informed that they would see sequences of three hexagons on the screen, and were asked to indicate the “odd-one-out” hexagon, namely the one that differed from the others. They were instructed to press the following buttons on the keyboard as fast and as accurately as possible: “1” when the “odd-one-out” stimulus appeared on the left side of the sequence, “2” when it appeared in the middle, and “3” when it appeared on the right side. As before, participants were instructed to place their fingers on the relevant keys in advance, in order to reduce cognitive load. The task started with a practice phase of six items. The main experimental phase consisted of four blocks with 30 items each

(5 items x 3 positions (right, left, middle) x 2 conditions (spatial, gap), resulting in a total of 120 experimental items). Item and block order were fully randomised. Figure 5.9 depicts a trial structure: for each trial, a fixation cross was presented on the screen for 2000 ms. The fixation cross was followed by the presentation of a stimulus trial. The stimuli remained on the screen for 100 ms and were immediately followed by a blank screen. The blank screen remained until participants pressed one of the three valid buttons (“1”, “2”, or “3”) or until time-out (5000 ms). The task lasted approximately 5 minutes.

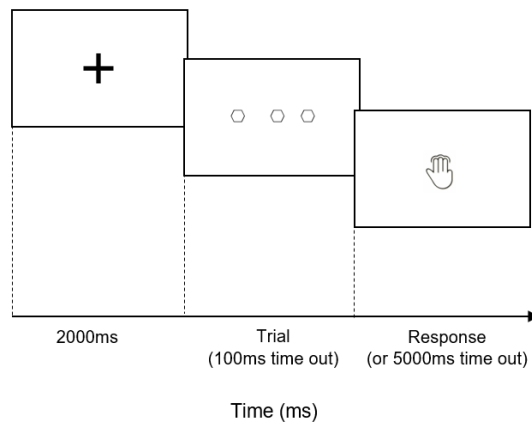


Figure 5.9: Schematic illustration of an experimental trial in the odd-man-out task.

5.2.4.2.3 Digit span task

Lastly, to measure WMC, a modified version of a digit span task was conducted, based on the WAIS-R Digit Span test (Wechsler 1987). For this modified version, stimuli recorded for a different experiment (Röhr, Savino, et al. 2022, Grice et al. 2024) were used. Participants were instructed to wear headphones as they would be presented auditorily with sequences that progressively increased in length, containing digits from 1 to 9. After every two sequences of the same length, starting with a three-digit sequence, the length of the next two sequences was extended by one digit, up to two sequences of nine digits, resulting in a total of 14 experimental sequences. Participants were asked to recall all digits of a sequence immediately after the presentation of the last digit, in the exact same order they were presented. Participants could hear each sequence only once. After sequence

5 Indexing attention orienting of domain-final rises using pupillometry

presentation, a numeric keypad appeared on the screen. Responses were entered by clicking each digit on the numeric keypad in the appropriate order. No digit could be omitted from the response. At the top left of the screen, above the keypad, a counter showed how many digits had already been entered. As soon as the participants entered the last digit of a sequence, they could proceed to the next sequence by clicking a “next” button. All stimuli were initiated by a beep tone of 890 ms, followed by 500 ms of silence, before the digit sequence started. The task lasted approximately 5 minutes. Figure 5.10 illustrates a trial structure.

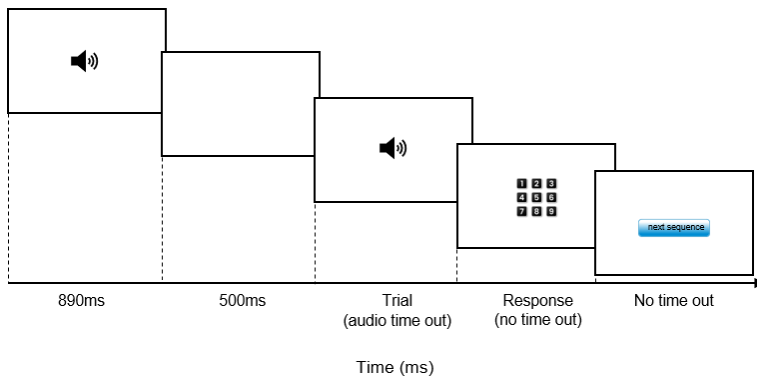


Figure 5.10: Schematic illustration of an experimental trial in the digit span task.

5.2.5 Data processing and statistical analyses

Data processing and statistical analyses were conducted in R, version 4.1.2 (R Core Team 2023), using the R packages *ggplot2* 3.3.5 (Wickham 2016), *itsadug* 2.4.1. (van Rij et al. 2022), *mgcv* 1.9-1 (Wood 2023), *PupilPre* 0.6.2. (Kyröläinen et al. 2020), and *tidyverse* 1.3.1 (Wickham et al. 2019). For reproducibility, data and scripts have been made available at <https://osf.io/nhy4c>.

5.2.6 Data pre-processing

5.2.6.1 Pupil data

Pupil data were exported using SR Research Data Viewer (v.4.3.210), and were further processed using the R-package *PupilPre* (Kyröläinen et al. 2020). Pupil data were re-aligned to 100 ms prior to the onset of the deviant number, and then continued for 3000 ms. Blink components were automatically detected and removed from the raw pupil data. The data were then manually checked, and

the remaining blink artifacts were removed by hand. Subsequently, trials including more than 20% of missing data because of blink artifacts were completely removed from further analyses, yielding 3.52% loss of the total dataset. After artifact rejection, the raw data were interpolated using cubic spline interpolation and then filtered with a Butterworth 0.1 Hz low-pass filter. Skipped trials due to missing values and artifacts created by the filter were removed using the *trim_filtered* function. Thereafter, the raw data were baseline-normalised by trial (subtractive correction) using the average of 100 ms preceding the onset of the deviant. Finally, normalised data were down-sampled to a rate of 100 Hz (10 ms time bins).

5.2.6.2 Cognitive test battery

For the flanker and odd-man-out tasks, accuracy (correct/incorrect responses, coded as 1/0) as well as response times (in ms) were recorded for each trial. To measure participants' inhibitory ability, the efficiency measure by Spillsbury et al. (1990) was employed. Specifically, the number of the correct incongruent trials was divided by the median response time (inhibition score = [number of correct answers in incongruent trials] / [median response time]) per participant. To measure processing speed, a similar efficiency measure was calculated per participant. This measure was computed by dividing the number of correct trials by the median response time. Lastly, in the digit span task, digit responses were recorded in the recalled order by participants. The digit span of each participant was calculated as the length of the last correctly recalled sequence before that participant's failure on two consecutive sequences.

Table 5.2 presents individual cognitive profiles in terms of inhibition, processing speed, and WMC. The table shows raw mean scores per individual participant across the three cognitive tasks implemented. Higher scores in the flanker task indicate better inhibitory skills; higher scores in the odd-man-out task indicate slower processing speed; and higher scores in the digit span task indicate larger WMC.

Table 5.2: Mean performance per participant across the three cognitive tasks.

Participant ID	Flanker	Odd-man-out	Digit Span
01	0.98	3.77	8
02	0.98	5.97	6
03	0.92	5.08	9

5 Indexing attention orienting of domain-final rises using pupillometry

Participant ID	Flanker	Odd-man-out	Digit Span
04	0.98	4.15	6
05	0.98	3.30	6
06	0.98	7.00	5
08	0.90	4.06	6
09	0.98	7.14	5
10	1.00	6.26	7
11	0.92	4.47	5
12	0.96	2.89	7
13	0.98	3.14	6
14	0.96	5.41	7
15	0.98	4.45	6
17	0.98	4.01	7
18	0.94	5.18	6
19	0.96	5.71	7
20	0.98	5.53	9
21	1.00	5.80	7
22	0.96	4.87	8
23	0.98	4.21	6
24	1.00	5.71	5
25	0.98	4.14	6
26	1.00	5.01	6
27	0.94	5.27	6
30	0.98	4.90	5
31	0.85	3.71	8
32	0.98	5.98	6
33	0.98	5.54	6
34	1.00	2.96	6
35	1.00	4.64	7
36	0.94	4.86	6
37	0.96	4.43	8
38	0.96	3.23	5
39	0.96	3.55	7
40	0.90	5.85	6
41	0.98	5.14	7
42	1.00	3.92	5
44	0.96	4.79	6
45	0.98	4.01	5

Participant ID	Flanker	Odd-man-out	Digit Span
46	1.00	8.05	6
47	0.98	4.97	4
48	0.92	5.12	7
49	0.94	3.83	6
50	0.92	4.39	8
51	0.96	7.72	7
52	1.00	4.10	5
53	0.90	3.76	6
54	1.00	6.46	6
55	1.00	4.95	6
57	0.98	4.98	6
59	0.96	2.94	8
60	0.96	4.56	7
61	0.92	5.49	6
62	0.88	5.04	7
63	0.98	5.30	5
64	1.00	4.71	8
65	0.98	3.76	6
66	0.96	5.28	9
67	0.92	3.58	6

The next processing step consisted of standardising (z-scoring) participants' scores across the flanker, odd-man-out, and digit span tasks. Although participants could be classified on the basis of their performance on one of the three cognitive tasks, the combinations of the skills measured in the three tasks is considered crucial for a better understanding of the different cognitive pathways involved in the process of OR activation. Prior to statistical modelling, correlation tests among the three cognitive scores were employed. As in the pilot study using the button press paradigm (see Section A.5.2), the correlation tests revealed a weak positive correlation between odd-man-out and flanker scores ($r(572278) = 0.34$, $p < 0.0001$), indicating that the slower the processing speed, the better the inhibitory skills (and vice versa). A weak negative correlation between digit span and flanker scores was also found ($r(572278) = -0.23$, $p < 0.0001$), indicating that the greater the span, the lower the inhibitory skills.

5.2.6.3 Inference criteria

The statistical analysis of the pupil data is divided into two parts, *confirmatory* and *exploratory*, similarly to the evaluation of the pilot results (see Section A.4). The confirmatory analysis tested the prediction that deviant numbers produced with rising intonation, due to its attention orienting function, will induce a stronger PDR compared to numbers realised with falling intonation, or when the intonation does not differ from that of the standard numbers, that is, deviants produced with neutral intonation.

In the exploratory analyses, I examined whether sequence length affects PDR and/or whether it interacts with the different prosodic deviant realisations. Besides sequence length, I also explored whether individual cognitive profiles interact with the prosodic conditions and thus affect the processing of the introduced deviances produced with different intonational patterns. The full model specifications can be found in the script provided on OSF.

The PDR was normalised and modelled using Generalised Additive Mixed Modelling (GAMM), which has been shown to be effective for analysing pupil data, as it accounts for nonlinear patterns and interactions, nonlinear random effects, and the inherent autocorrelation of time-course data (see van Rij et al. 2019). In particular, GAMM is appropriate for modelling nonlinear time-series patterns, capturing variation in two trajectories: *height* and *shape*. These two trajectories are captured by different terms: parametric terms allow for mean differences in the overall height of the curves, and smooth terms allow for differences in the shape of the curves. GAMM also accounts for random effect structures by using random smooths (e.g., Winter & Wieling 2016, Wood 2017, Sóskuthy 2017). Random smooths expand the principle of smooth terms to the random effects by fitting separate smooths at each value of a grouping variable, thereby allowing different curve shapes for different subjects and/or items.

5.2.6.4 Confirmatory analysis

PDR was modelled as a function of the ordered factor⁶ PROSODIC CONDITION. Treatment contrast was used to code PROSODIC CONDITION (levels: rise/fall/neutral), with rise serving as the reference level. This coding allows for testing the following contrasts:

⁶Ordered factors allow for testing whether the curves of each level of the factor differ not only in height (parametric coefficients) but also in shape (difference smooth terms).

1. rise vs. fall, and
2. rise vs. neutral.

The model included PROSODIC CONDITION both as a parametric term, testing for overall *height* differences in PDR curves between prosodic conditions, and as a smooth reference term, capturing the *shape* effects in the reference level of the PROSODIC CONDITION (i.e., rise) over TIME. The model also included a difference smooth term by PROSODIC CONDITION, testing *shape* differences of PDR curves between prosodic conditions (i.e., rise vs. fall, and rise vs. neutral) over TIME. Further, the model included a random smooth by SUBJECT, and a reference-difference random smooth for each SUBJECT by PROSODIC CONDITION which captures *shape* differences by SUBJECT. Lastly, autocorrelation within trajectories was controlled via the inclusion of an AR1 residual model. Given that this model could not test for the contrast between fall vs. neutral, the simultaneous confidence interval (CI) test⁷ was implemented to examine this contrast.

5.2.6.5 Exploratory analyses

Moving to the exploratory part of this study, for both *sequence length* and *individual cognitive variability*, I fitted three separate models to examine their interaction with PROSODIC CONDITION in the three following contrasts:

1. *Model 1*: neutral vs. rise,
2. *Model 2*: neutral vs. fall, and
3. *Model 3*: rise vs. fall.

Results are presented on the basis of model summaries.

5.2.6.5.1 Sequence length

For the investigation of the *sequence length* effect, PDR was modelled as a function of the ordered factors PROSODIC CONDITION (treatment coded) and LENGTH (treatment coded, reference level: long), as well as their interaction. PROSODIC CONDITION, LENGTH, and their interaction were included in the models as parametric terms and smooths over TIME. In addition, random smooths by SUBJECT,

⁷The simultaneous confidence interval (CI) test can be used as a proxy for a post-hoc test: when testing two whole curves simultaneously, if any point in the CI does not include zero, then the difference between them can be treated as significant.

and reference-difference random smooths per condition for the individual levels of SUBJECT, were included in all models. Lastly, an AR1 residual model was included in all models to control for autocorrelation within trajectories. Model 1 was fitted such that the smooth over TIME for the condition [neutral, long] represented the reference smooth. The model further included difference smooths for the conditions [rise, long], [neutral, medium], and [rise, medium]. This approach allows for assessing the following differences: the difference smooth for [rise, long] reflects the effect of rising vs. neutral deviants in long sequences; the difference smooth [neutral, medium] shows the effect of neutral deviants in medium-length vs. long sequences; and the difference smooth [rise, medium] highlights the effect of rising deviants in medium-length vs. long sequences. Model 2 was fitted such that the smooth over TIME for the condition [neutral, long] represented the reference smooth. The model further included difference smooths for the conditions [fall, long], [neutral, medium], and [fall, medium]. Specifically, the difference smooth for [fall, long] reflects the effect of falling vs. neutral deviants in long sequences; the difference smooth [neutral, medium] shows the effect of neutral deviants in medium-length vs. long sequences; and the difference smooth [fall, medium] highlights the effect of falling deviants in medium-length vs. long sequences. Finally, model 3 was fitted such that the smooth over TIME for the condition [rise, long] represented the reference smooth. The model further included difference smooths for the conditions [fall, long], [rise, medium], and [fall, medium]. Specifically, the difference smooth for [fall, long] reflects the effect of falling vs. rising deviants in long sequences; the difference smooth [rise, medium] shows the effect of rising deviants in medium-length vs. long sequences; and the difference smooth [fall, medium] highlights the effect of falling deviants in medium-length vs. long sequences.⁸

5.2.6.5.2 Cognitive variability

For the exploration of the *individual cognitive variability* effect, and given the aforementioned correlations among the cognitive predictors, only one predictor was selected for the subsequent analyses. Due to its high relevance for attention orienting, the *flanker scores*, indicating inhibitory skills, were chosen. Therefore, PDR was modelled as a function of the ordered factor PROSODIC CONDITION (treatment coded) and its interaction with the continuous FLANKER SCORES. Specifically, PROSODIC CONDITION was included in the models as a parametric term, with

⁸These models do not allow for estimating the following comparisons: [rise, medium] vs. [neutral, medium]; [fall, medium] vs. [neutral, medium]; [rise, medium] vs. [fall, medium]. Given that these differences are not of primary interest, I describe them based on the illustration of GAMM smooths.

a reference smooth over TIME and a difference smooth over TIME by PROSODIC CONDITION. Further, the models included a reference and a difference smooth over FLANKER SCORES by PROSODIC CONDITION. The reference smooth captured non-linear changes (if any) in the average PDR as a function of the FLANKER SCORES for the reference prosodic level (neutral in model 1; falling in model 2; and rising in model 3), while the difference smooth reflected shape differences in the non-reference prosodic level (rising in model 1; neutral in model 2; falling in model 3), compared to the reference level. A tensor product reference smooth and a tensor product difference smooth were also parts of the model, capturing the two-way interaction between PROSODIC CONDITION and FLANKER SCORES over TIME. The tensor product reference smooth modelled the effect of FLANKER SCORES (if any) on the curve shape of the prosodic reference level (neutral in model 1; falling in model 2; and rising in model 3). The tensor product difference smooth reflected whether this effect changed in the non-reference prosodic level (rising in model 1; neutral in model 2; and falling in model 3). In addition, random smooths by SUBJECT and reference-difference random smooths per condition for the individual levels of SUBJECT were included in all models. Lastly, an AR1 residual model was included in all models to control for autocorrelation within trajectories.

5.3 Results

This section presents results reporting model predictions and plotting GAMM smooths which depict the different estimated effects on the PDR within the time window of interest. The presentation of the results is structured as follows. First, I concentrate on the confirmatory analysis, that is, I report on PDR differences among prosodic conditions, and thus focus on results related to the hypothesis that rising deviants may induce stronger PDRs due to the attention orienting function of rising intonation. Subsequently, I report on the exploratory analyses, presenting the effects of sequence length and individual cognitive variability on PDRs, and hence on the processing of the introduced deviances produced with different intonational patterns.

Figure 5.11 illustrates grand averaged changes in pupil diameter from the baseline average over time (sampled in 10 ms bins), as a function of prosodic condition, time-locked to the onset of the deviant stimulus (zero ms; as depicted by the vertical dashed line). The yellow point-range curve depicts pupillary responses to deviants produced with the rising edge tone (rising condition), the green point-range curve shows pupillary responses to deviants featuring the

5 Indexing attention orienting of domain-final rises using pupillometry

falling edge tone (falling condition), and the black point-range curve illustrates pupillary responses to deviants with neutral intonation (the baseline prosodic condition). Visual inspection of the curves reveals that deviant numbers elicit a PDR regardless of the intonation they feature, but the different intonational contours seemed to affect the PDR in distinct ways.

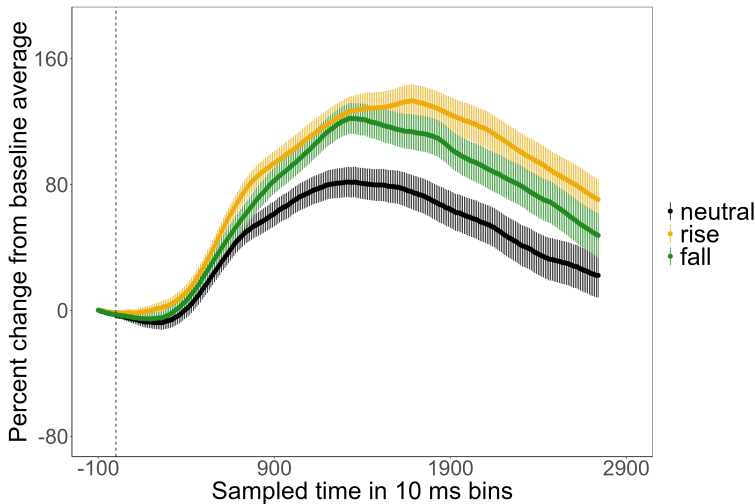


Figure 5.11: Grand averaged pupil dilation response per prosodic condition over time (sampled in 10 ms bins), time-locked to the onset of deviant stimulus (illustrated by the vertical dashed line), and estimated from the baseline average.

5.3.1 Confirmatory analysis

GAMM smooths⁹ (including 95% CIs) of PDR to deviant numbers as a function of prosodic condition are depicted in Figure 5.12. Colour coding of the prosodic conditions' estimated smooths corresponds to the colour coding in Figure 5.11.

Comparing PDR curves between rising and neutral conditions, the model revealed that rising intonation elicited a more robust PDR, in terms of both the *overall height* (parametric difference: $\beta = -37.06$, $t = -4.380$, $p < 0.001$) as well as the *shape* of the PDR curve (smooth difference: EDF = 3.198, $F = 7.038$, $p < 0.001$), reflecting an increased and long-lasting effect. For the contrast between falling and neutral intonation, the simultaneous confidence interval test indicates that the PDR curve associated with falling edge tones was significantly different from

⁹GAMM smooths illustrate height, shape, and latency properties of the effects.

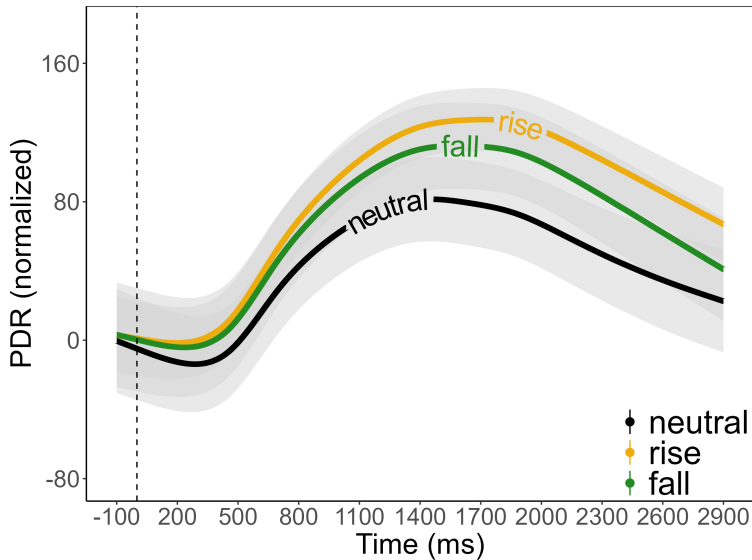


Figure 5.12: GAMM smooths of PDR (normalised) with 95% CIs, across prosodic conditions. PDRs are time-locked to the onset of the deviant (zero ms), as indicated by the vertical dashed line.

the PDR curve associated with neutral intonation ($t = 2.357$), in that the amplitude is larger and prolonged. Finally, when comparing rising to falling intonation, rises are differentiated from falls by the *shape* of the PDR curve (smooth difference: EDF = 1.048, $F = 3.632$, $p = 0.05$), meaning that rising intonation exhibits a more sustained effect over time.

In sum, both rising and falling edge tones evoked greater pupillary dilations, both in magnitude and duration, than the baseline neutral intonation. At the same time, PDR to rises showed a more sustained effect over time compared to the falling condition.

5.3.2 Exploratory analyses

5.3.2.1 Sequence length

GAMM smooths (including 95% CIs) of PDR to deviant numbers, as a function of prosodic condition and sequence length are shown in Figure 5.13. As before, black depicts the estimated smooths of deviants neutrally produced, yellow depicts the estimated smooths of deviants produced with rising intonation, and green depicts the estimated smooths of deviants produced with falling intonation. Solid and dashed lines represent long and medium-length sequences, respectively.

5 Indexing attention orienting of domain-final rises using pupillometry

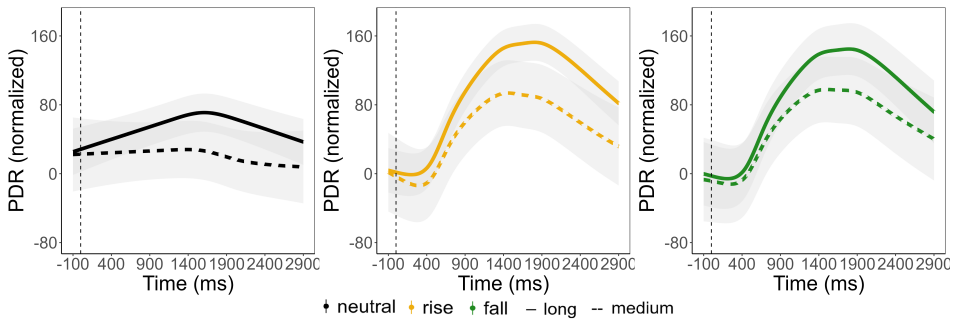


Figure 5.13: GAMM smooths of PDR (normalised) with 95% CIs, across prosodic conditions and sequence lengths. PDRs time-locked to the onset of the deviant (zero ms), as indicated by the vertical dashed line.

The models revealed an interaction between prosodic condition and sequence length. Specifically, when comparing rising to neutral deviants (see solid lines in middle and left panels of Figure 5.13), and falling to neutral deviants (see solid lines in right and left panels of Figure 5.13) in long sequences, the results showed that both rising and falling deviants elicited greater PDRs in terms of both the *overall height* ([rise vs. neutral] parametric difference: $\beta = 42.96$, $t = 3.259$, $p = 0.001$; [fall vs. neutral] parametric difference: $\beta = 35.35$, $t = 2.904$, $p = 0.001$) and *shape* of the PDR curve ([rise vs. neutral] smooth difference: EDF = 5.375, $F = 5.602$, $p < 0.001$; [fall vs. neutral] smooth difference: EDF = 4.291, $F = 5.699$, $p < 0.001$), demonstrating an increased and long-lasting effect. Nevertheless, no differences were found between rising and falling deviants in long sequences (see solid lines in middle and right panels of Figure 5.13; [rise vs. fall] parametric difference: $\beta = -6.984$, $t = -0.622$, $p = 0.53$; smooth difference: EDF = 1.011, $F = 0.103$, $p = 0.76$).¹⁰

When comparing rising and falling deviants in long sequences to rising and falling deviants in medium-length sequences (see middle and right panels of Figure 5.13, respectively), the models show that both edge tones evoked PDRs of smaller amplitude and shorter latency in medium-length sequences as opposed to long sequences ([rise, long vs. rise, medium] parametric difference: $\beta = -32.20$,

¹⁰Although the current models could not test for the same comparisons in medium-length sequences, it can be seen from the GAMM smooths in Figure 5.13 that the differences in medium-length sequences follow the same trends as in long sequences. That is, on the one hand both rising and falling deviants in medium-length sequences appear to evoke more robust PDRs than neutral deviants (see dashed lines in Figure 5.13), as they do in long sequences. On the other hand, it appears that rising deviants do not differ from falling deviants in medium-length sequences, similarly to the long sequences.

$t = -2.747$, $p = 0.001$; smooth difference: EDF = 4.514, $F = 3.210$, $p = 0.001$; [fall, long vs. fall, medium] parametric difference: $\beta = -43.86$, $t = -2.795$, $p = 0.001$; smooth difference: EDF = 3.271, $F = 5.017$, $p < 0.001$). However, when comparing neutral deviants in long and medium-length sequences (see left panel of Figure 5.13), PDRs to medium-length sequences differ only in *shape*, in that they exhibit a less curvy line, indicating a faster response ([neutral, long vs. neutral, medium] smooth difference: EDF = 1.034, $F = 5.380$, $p = 0.01$), while no *height* differences are observed ([neutral, long vs. neutral, medium] parametric difference: $\beta = -17.86$, $t = -1.235$, $p = 0.21$).

In sum, rising and falling edge tones evoked PDRs of greater magnitude and latency than the baseline neutral condition, in long (and medium-length) sequences. Within both rising and falling edge tones, the long sequences evoked PDRs of larger amplitude and prolonged latency as opposed to medium-length sequences. This was not the case for neutral intonation. For this prosodic condition, long sequences elicited longer pupillary responses, which, however, did not differ in amplitude compared to those elicited by medium-length sequences.

5.3.2.2 Cognitive variability

GAMM smooths (including 95% CIs) of PDR to deviant numbers, as a function of prosodic condition and flanker scores, are shown in Figure 5.14. Flanker scores are illustrated from left (*lower inhibition*) to right (*higher inhibition*) panels. Figure 5.15 illustrates difference smooths (calculated as smooth A – smooth B) with 95% CIs for the comparisons among prosodic conditions as a function of flanker scores (x-axis). The red shaded areas depict the points in the 95% CI that do not include zero, which are regarded as the windows of significant differences.

The models indicated an interaction between prosodic condition and flanker scores. More specifically, the results show that PDR changes as a function of inhibitory ability across prosodic conditions. For neutral intonation (horizontal top [and middle] panels in Figure 5.14), it was shown that inhibitory ability modulated PDR *shape*, such that the lower the flanker score, the longer the PDR (tensor product smooth: EDF = 8.442, $F = 2.758$, $p = 0.001$). Similarly, for falling intonation (horizontal middle [and bottom] panels in Figure 5.14), inhibitory ability also affected PDR *shape*. From low to average to high flanker scores, a decrease in the duration of the PDR was observed (smooth difference: EDF = 1.900, $F = 4.657$, $p = 0.01$; tensor product smooth: EDF = 10.322, $F = 4.581$, $p < 0.001$). For rising intonation (horizontal [top and] bottom panels in Figure 5.14), the results also show that PDR *shape* changed as a function of inhibitory ability, such that the better the flanker score, the longer the PDR (tensor product smooth: EDF = 8.866, $F = 4.476$, $p < 0.001$).

5 Indexing attention orienting of domain-final rises using pupillometry

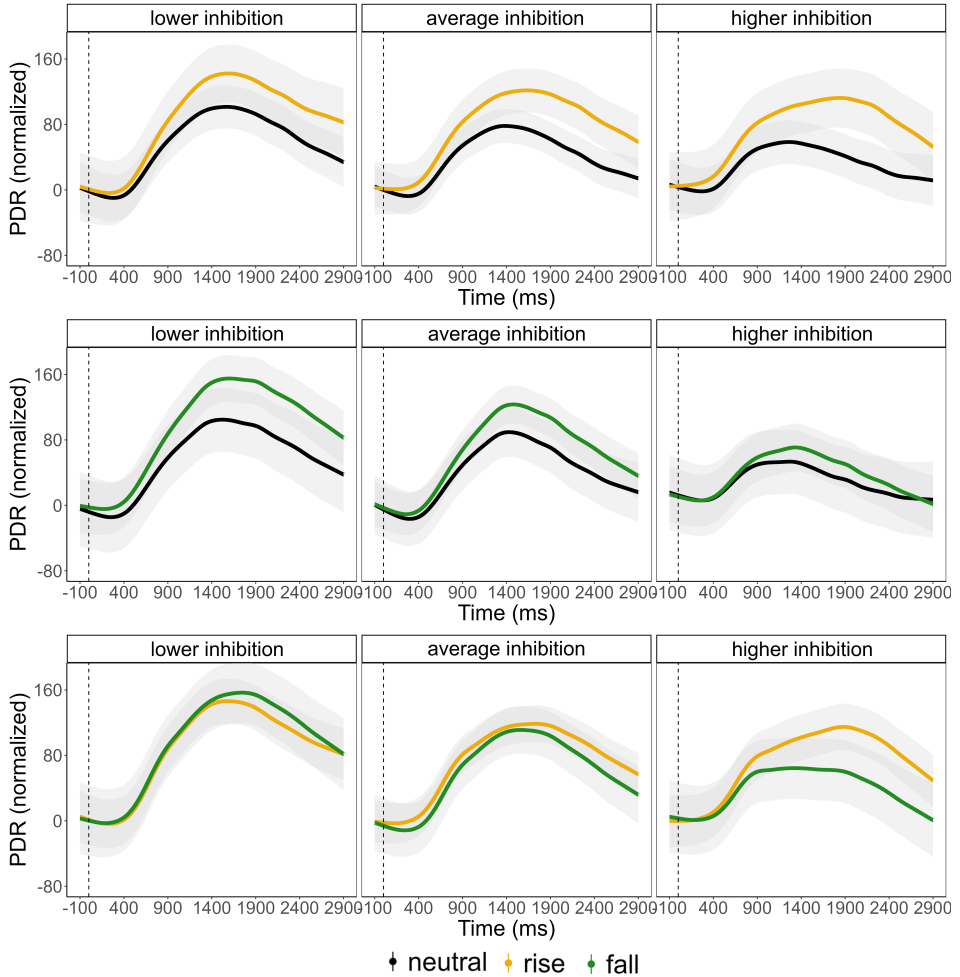


Figure 5.14: GAMM smooths of PDR (normalised) with 95% CIs, across prosodic conditions and flanker scores. PDRs are time-locked to the onset of the deviant (zero ms), as indicated by the vertical dashed line. Yellow depicts the estimated smooths of deviants produced with rising intonation, green depicts the estimated smooths of deviants produced with falling intonation, and black depicts deviants produced neutrally.

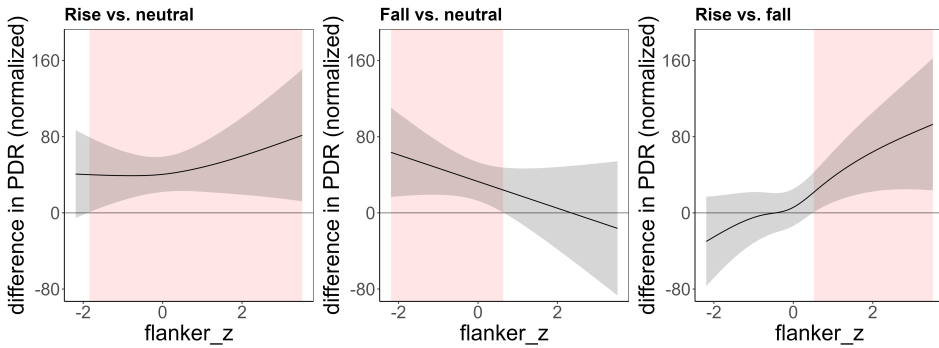


Figure 5.15: Difference smooths of PDR (normalised) with 95% CIs for the comparisons among prosodic conditions as a function of flanker scores. The red shaded areas depict where the 95% CI does not include zero, indicating the windows of significant differences.

Comparing neutral to rising intonation (horizontal top panels in Figure 5.14 and left panel in Figure 5.15), PDRs differed in both *height* and *shape*, as a function of inhibitory skills: from low to high flanker scores, the higher the score, the stronger the rising PDRs, showing an increased and long-lasting effect (parametric difference: $\beta = 37.471$, $t = 4.294$, $p < 0.001$; smooth difference EDF = 5.453, $F = 6.164$, $p < 0.001$; tensor product smooth difference: EDF = 6.162, $F = 3.346$, $p = 0.001$). Likewise, comparing falling to neutral intonation (horizontal middle panels in Figure 5.14 and middle panel in Figure 5.15), PDRs differed in both *height* and *shape* as a function of low inhibitory skills, such that the lower the flanker scores, the weaker the neutral PDRs, showing a decreased and subtly faster effect (parametric difference: $\beta = -23.014$, $t = -2.337$, $p = 0.01$; smooth difference EDF = 5.170, $F = 3.422$, $p = 0.001$; tensor product smooth difference: EDF = 1.003, $F = 2.757$, $p = 0.05$). Finally, comparing rising to falling intonation (horizontal bottom panels in Figure 5.14 and right panel in Figure 5.15), PDRs differed only in *shape* as a function of high inhibitory skills, such that the higher the flanker score, the faster the falling PDRs (smooth difference: EDF = 1.898, $F = 2.427$, $p = 0.05$; tensor product smooth difference: EDF = 7.762, $F = 2.887$, $p = 0.001$).

In sum, individuals' inhibitory skills modulated PDR to deviants differently across prosodic conditions. Individuals with higher flanker scores, and thus better inhibitory skills, showed sustained PDRs only to rising deviants, as opposed to neutral and falling ones, with no difference between the latter two ([rise] > [fall = neutral]). In contrast, individuals with lower flanker scores, and hence weaker inhibitory skills, exhibited sustained PDRs to both rising and falling PDRs, but not to the baseline neutral ([rise = fall > neutral]).

5.3.3 Interim summary

To briefly summarise the current findings, deviants produced with rising and falling intonation (high and low edge tones, respectively) elicit a greater pupillary response (both in magnitude and duration) than deviants featuring the baseline neutral intonation. This holds for both long and medium numeric sequences. Within the edge tone conditions, rising intonation resulted in a greater pupillary response than falling intonation, manifesting as a more sustained effect over time. In the next section, I will argue that the gradient effects of intonation on PDR response indicate different modulations of attention orienting in each prosodic case.

Turning to the sequence length effect, the results show that both rising and falling edge tones evoked stronger PDRs, showing an increased and long-lasting effect in long sequences compared to medium-length sequences. However, neutral intonation elicited only subtly longer pupillary responses in long sequences, which, in turn, did not differ in height amplitude compared to those elicited in medium-length sequences. In the next section, I will reason that the findings we observe here reflect the temporal conditions under which our auditory processing system develops sensory-memory traces of the stimuli and their regularities, which are crucial representations for the mechanism which underlies the attention orienting response.

Finally, considering individual variability, it is evident that individuals' inhibitory skills affect PDR modulation across prosodic conditions. For individuals with strong inhibitory skills, only rising intonation led to more sustained PDRs, while falling and neutral ones evoked transient pupillary responses. In contrast, for individuals with weaker inhibitory skills, both rising and falling prosodic conditions caused prolonged PDRs, as both led to stronger responses than the neutral baseline. In the following, I will argue that the observed PDR responses in relation to inhibition indicate that the successful evaluation of the deviant, and thus the activation of the different stages of the orienting response, depends on individual-specific cognitive bandwidth.

5.4 Discussion

In this chapter, I investigated the relevance of domain-final rises and falls (the reflexes of phrase final edge tones) for attention orienting in German. Domain-final rising intonation has been shown to attract attention in serial recall experiments (e.g., Savino et al. 2020, Röhr, Savino, et al. 2022, Grice et al. 2024). Here, I tested the attention orienting function of domain-final rises in a different way, namely

by utilising a changing-state oddball paradigm in a pupillometry study. The aim of the study was trifold. Using PDR as a reflex of attention orienting, the main objective was to test whether deviant numbers in seriatim ascending sequences can capture more attention, thereby evoking more robust PDRs, when realised with a final rise compared to when they are produced with a final fall or when their intonation does not differ from that of the standards (the neutral case). In addition, this study took an explorative direction in the examination of sequence length effect on PDR elicitation, and the contribution of individual cognitive variability to the processing of deviances, with the aim of better understanding the function of attention orienting.

Prior to the pupillometric study, the design of the stimuli was piloted using a button press paradigm in an explicit deviant-detection task (see Appendix A). It was confirmed that the design of the stimuli was successful, in that listeners indeed perceived and detected the deviances across numeric sequences. The pilot data provided some further interesting results which are discussed in Section A.6. The structure of the current section is the following. First, I put the findings in the broader context of intonational rises vs. falls and their relevance for attention orienting, and consequently, I elaborate on the implications of these findings for the function of rising intonation at the edges of constituents as an attention orienting device. Next, I highlight the relevance of sequence length for the orienting response in passive listening. Finally, I discuss the contribution of individual cognitive variability in the activation of attentional mechanisms.

5.4.1 Rising intonation at the edges of constituents as an orienting device

Starting with the group level findings from the confirmatory part of this study, my predictions were the following. First, given that attention orienting is indexed by dilations in pupil size (e.g., Liao et al. 2016, Marois et al. 2018, 2019), and given the claim that attention orienting is underlined by a mechanism driven by expectancy violations (e.g., Näätänen et al. 2011, Vachon et al. 2012), I predicted that the presentation of a deviant number in a seriatim ascending numeric sequence would disrupt the anticipated pattern, causing an attention switch expressed in a PDR. In addition, given the association between attention orienting and rising pitch reported in previous work (e.g., Näätänen et al. 1978, 1980, Alain et al. 1994, Doeller et al. 2003, Chobert et al. 2012, Hsu et al. 2015, Ventura et al. 2020, Röhr et al. 2021, Lialiou, Grice, et al. 2024), the second prediction was that deviants produced with the rising edge tone would result in greater disruption, expressed in a more robust PDR compared to deviants produced with the falling edge tone or with neutral intonation.

The present pupillometric results provide evidence in favour of the first prediction: deviant numbers do indeed elicit a PDR, regardless of intonation. These results corroborate previous findings on auditory cognition claiming that attention orienting is underpinned by an expectancy violation mechanism (e.g., Hughes et al. 2007, Vachon et al. 2012, Paavilainen 2013, Hughes 2014). Now, when the intonation of deviants differed from the intonation of standards, exhibiting either rising or falling edge tones, deviants elicited an increased pupil dilation effect compared to the neutral condition, in which the intonation of the deviant did not differ from that of the standards. This finding indicates that prosodic marking on deviants results in more robust attentional resources being allocated towards the violation. This finding is compatible not only with the idea from the early literature that attention orienting is sensitive to the physical properties of auditory deviances (for more, see Wright & Ward 2008), but also with results from neurocognitive studies reporting that signals with both rising and falling acoustic properties attract attention when presented as deviants (for literature review, see Section 2.4).

Moving to the second prediction, and hence the comparison between rising and falling edge tones, a subtle PDR *shape* difference was found between the two tones. This difference indicates that rising tones evoked more long-lasting changes to PDRs than falling tones. These results potentially reflect qualitatively different cognitive processes. As discussed in Section 2.5, the latency of attention-related PDRs is indicative of the processing stage in which the attentional mechanism is activated: transient dilations reflect pre-attentive attentional processes, while more prolonged responses indicate processes at the conscious level (see Strauch et al. 2022). As discussed in Chapter 2, attention-related PDRs consist of different subcomponents, two of which are the orienting-related and the executive-related PDR. Orienting-related PDRs are usually momentary, and typically evoked by sensory events such as an instance of auditory deviance, while executive-related PDRs are prolonged and evoked by top-down processes. Thus, an orienting-related PDR, evoked by an auditory deviant, reflects an involuntary attention switch, which in turn, and after some evaluation of the deviant in the processing system, can further activate the executive network, thereby bringing the deviant into awareness. In other words, the route from pre-attentive to conscious perception of a deviant is reflected in the time course of pupillary responses. The findings of this study indicate that, on the one hand, falling edge tones induce an involuntary attention switch towards the deviant pre-attentively, as reflected in the observed transient pupil dilation. On the other hand, rising edge tones initially cause a similar involuntary switch of attention, and then, potentially due to their prominence, bring the perception of the deviant into

awareness, thereby activating voluntary attention orienting. This finding can be related to previous neurocognitive studies (e.g., Rinne et al. 2005, Macdonald & Campbell 2011) suggesting a differential processing of rises and falls. Whilst deviant sounds with both rising and falling acoustic properties have been found to attract attention in auditory cognition, rises have been claimed to serve as intrinsic warning cues due to their saliency (e.g., Bach et al. 2008). Specifically, rises have been reported to prompt auditory looming effects, which indicate an approaching sound source, whereas falls are experienced as receding sound sources, activating fewer attentional resources than rises (for literature review, see Section 2.4). Let me now elaborate on how the results of this study elucidate the aforementioned point in the linguistic domain.

In the two oddball conditions discussed above, the deviant numbers are realised with rising and falling edge tones, whereas their corresponding standard stimulation is produced with neutral intonation (shallow falling intonation; for the acoustic characterisation of the intonational contours see Section 5.2.3). In such sequences, the listener potentially builds predictions both on the basis of the sensory input and on the linguistic meaning this sensory input transmits. First, listeners detect regularities in the sensory input, then they extrapolate a pattern, and finally, they predict upcoming events (for more on the auditory system operations, see Section 2.3). Thus, the listener, after being exposed to the seriatim ascending numeric pattern produced with neutral intonation, not only predicts and anticipates the following stepwise ascending number, but also anticipates that this number will be produced with the same intonational pattern. Now, as shown in Chapter 4, predictions can also arise from linguistic meaning. The items used in this study are numeric sequences, which inherently give rise to expectations about forthcoming numbers. On top of this natural prediction context created by the (semantics of the) numeric sequences, the standard/seriatim numbers are produced with the aforementioned neutral intonational contour, a contour which can be featured on non-final items of a sequence or a list in German (for more on German list intonation, see Section 3.5). The expectations about a forthcoming number are therefore enhanced by the linguistic meaning of the intonation with which the standard numbers are realised, that is, a natural and appropriate contour on non-final list items which denotes that the sequence is not over yet. Now, the intonation of standard stimulation might have elicited additional meaning-based predictions, specifically, expectations about the forthcoming end of the numeric sequence. Therefore, the listener, on the one hand, anticipates a forthcoming ascending number, on the basis of both the seriatim pattern and the intonational meaning. On the other hand, the listener expects that the sequence will end at some point, and therefore seeks an intonational

cue that signals this end. Both rising and falling edge tones, although they denote different things (rises indicate continuity, whereas falls indicate finality), can mark the end (or the boundary) of a smaller (or larger) unit within the utterance. Thus, domain-final rises and falls mark the end of smaller units within sequences and are contextualised in a similar way in the items of this study: the deviant number, produced with either of the two tones, is the last item (but still part) of the same sequence unit – but not the last item of the entire list – thereby violating the anticipated seriatim pattern. Furthermore, the presentation of the edge tones violates the expectation that the next number should be produced with neutral intonation. This means that, on the one hand, rising and falling edge tones validate the anticipation of a forthcoming sequence end, and on the other hand, they violate the anticipated seriatim pattern produced with neutral intonation, since they mark a deviant number included in the same unit with the standard numbers.

The initial observation is that the presentation of a deviant, produced with either of the two edge tones, evokes an increased PDR (compared to a neutral deviant), indexing a greater switch of attentional resources towards the violation of the seriatim pattern produced with neutral intonation. Nonetheless, when the deviant number is produced with rising pitch, this elicited subtly prolonged pupil dilations, indicating that, in this case, the violation first causes an involuntary attention switch, followed by voluntary attention orienting and thus conscious processing. In a linguistic context where deviants produced with either of the two tones generate and violate the same signal-driven and meaning-based expectations, the ability of rising deviants to ultimately lead to conscious attention orientation, in contrast to falling (and neutral) deviants, can be attributed to their prominence. This is in line with Röhr et al. (2021) showing that the amount of attention oriented to a stimulus is defined by signal-based cues combined with meaning-based expectations derived from the context. This is also manifested in the results reported in Chapter 4, where I showed that involuntary attention orienting is signal-driven, while voluntary attention is driven by meaningful aspects of intonation licensed by contextually created expectations (see Section 4.4).

We can argue that the present results highlight the attention orienting function of rising pitch, suggesting that domain-final rises on deviant stimuli also enhance the ability of the deviant to attract attention at a more conscious level. This finding thus strengthens the case for the role of rising intonation at the edges of constituents in attention orienting. As discussed in Chapter 3, intonational events are phonologically anchored to specific positions in the prosodic structure, that is, they are either associated with the stressed syllable (pitch accents), or with the edges of constituents (edge tones). In the autosegmental-metrical

theory of intonational phonology (e.g., Ladd 2008) pitch accents are associated strictly with a prominence-cueing function, while edge tones are associated with a phrasing function. In that sense, it has been claimed that accentual rises are better in directing listener attention than rises at edges (for more, see Section 3.1). This has already been called into question by results from serial recall studies (e.g., Savino et al. 2020, Röhr, Savino, et al. 2022, Grice et al. 2024) which report that rising edge tones marking the final item of non-final triplets boost the recall accuracy of the whole triplet, hence orienting attention to the whole domain (see also Section 3.3). The work reported in Chapter 4 adds to those first indications of domain-final rising intonation attracting attention in showing that, during on-line processing, rising intonation takes on a special role in involuntary attention orienting, regardless of whether the rise played the role of accentual or boundary contour. Therefore, the results of the current study contribute further to this novel discovery that rises associated with the edges of constituents can also direct listener attention at a more conscious level. Building on the aforementioned serial recall results, one speculation is that domain-final rises attract more attention than falls because rising edge tones attract attention to the entire domain. It is hence possible that rising edge tones encourage the listener to more quickly integrate the deviant with previous items in the same domain, and that listener awareness of the violation thereby increases.

The privileged status of the rising edge tones is also shown in the pilot results on speed-accuracy trade-off and individual variability (see Section A.5.2). In the rising condition, no trade-off between speed and accuracy was observed, whereas such a trade-off was found in both the neutral and falling conditions. Specifically, faster responses in the neutral condition and slower responses in the falling condition lead to worse performance. Further, by taking into account the cognitive profiles of the listeners in the pilot study, we can see that listeners with better inhibitory skills were more accurate in identifying the rising deviants than the average, jolting them out of their inhibition. These results show that, even in a relatively easy explicit task where listeners are aware of the deviant numbers, neutral and falling intonation can be somewhat challenging for the listeners, whereas rising intonation appears to capture listener attention more readily.

5.4.2 The relevance of sequence length for attention orienting in language

Moving to the more exploratory part of this study, I first focus on the relevance of sequence length for attention orienting in language. The pilot data show no

difference between the two sequence lengths across all conditions, meaning that listeners were equally good and equally fast in detecting the deviant in both medium-length and long sequences, when explicitly asked to do so (see Section A.5.2). In contrast, the pupil results show that deviants in long sequences evoked more robust PDRs than deviants in the medium-length sequences across all prosodic conditions, with the neutral condition exhibiting this difference more subtly. The interpretation put forward here is that in passive listening, long sequences allow the auditory processing system to develop stronger memory traces, i.e., representations, of regularities in stimulus sequences, which are crucial for the expectancy violation mechanism. To illustrate the aforementioned point, consider how our sensory systems operate. According to Näätänen et al. (2011), the sensory systems develop representations, called memory traces, related to the stimuli and their regularities. The perception of a stimulus occurs at the transient phase of memory trace formation. Our brain continuously updates the representations in order to maintain a correspondence with the environment. During this updating process, the system extrapolates a pattern and forms certain predictions. When these predictions are not attested in the input that the system perceives, violating the representations formed, attention-calling mechanisms are activated (for more on the operations of the auditory system, see Section 2.3, and among others, Friston 2010, 2018, Näätänen et al. 2019).

Neurocognitive studies using simple sine waves have shown that for the memory trace of standard stimulation in passive auditory oddball paradigms to be formed, the presentation of standard auditory events, in the beginning as well as during the experimental session, needs to exceed a threshold of about 200 ms, also called the temporal window of integration (TWI; see Näätänen et al. 2019). The presentation of the standard numbers before the deviant in both sequence lengths of this study's items undoubtedly exceeded this TWI, simply because the items consisted of quadrisyllabic and pentasyllabic numbers, meaning that each number itself had a duration of 1000 ms or more. Both sequence lengths presented a sufficient amount of standard numbers before the deviant, specifically, 10 standard numbers (TWI of ca. 10 seconds) in the medium-length and 15 standard numbers (TWI of ca. 15 seconds) in the long condition, since deviants in both length sequences activated attentional mechanisms as reflected in the elicited PDR responses. The fact that PDRs were more robust in the long sequences shows that the longer TWI allows for a better and stronger formation of the memory trace corresponding to the standard stimulation, in turn leading to a greater switch of attention when this memory trace is violated. Therefore, these findings attest that in language, by virtue of the stimuli being of longer duration than the simple sine waves used in studies of auditory cognition, a TWI

of at least 10 seconds is needed to achieve the memory trace formation of standards and the concordant pattern extraction and prediction build-up. A longer TWI may help the processing system to build stronger memory traces, in turn leading to a more robust switch of attentional resources towards a violation. Nevertheless, for explicit deviant-detection tasks such as in the pilot study (see Appendix A), it appears that the longer sequences do not provide any advantage over the medium-length sequences.

5.4.3 The contribution of individual variability to attention orienting

A final, exploratory objective that this study put forward concerns the contribution of individual cognitive variability to attention orienting. Participants' individual variability was measured on the basis of three cognitive skills: processing speed, inhibitory ability, and WMC. The cognitive measurements yielded weak correlations (see Section 5.2.5), and thus the pupillary statistical analysis modelled only an interaction between flanker scores, reflecting participants' inhibitory ability, and prosodic conditions. The data show that the presentation of deviants attracts listeners' involuntary attentional resources, regardless of intonation. This is manifested in the dilations of participants' pupil size across all prosodic conditions. The data further indicate different PDR modulations for the three intonational patterns marking the deviants, as a function of inhibitory ability. More specifically, individuals with stronger inhibitory ability exhibited sustained PDRs to deviants produced with rising intonation, in contrast with more rapid PDRs to deviants produced with falling or baseline neutral intonation. Conversely, individuals with weaker inhibitory ability responded to both rising and falling deviant prosodic realisations with equally sustained PDRs, causing a more increased and prolonged effect than the neutral baseline.

As mentioned previously in this chapter, the time course of pupillary response is indicative of the different stages of the orienting response: transient dilations indicate pre-attentive processes; more prolonged dilations reflect processes at the conscious level. Therefore, the present data indicate that the presentation of the deviant first causes an involuntary attention switch, across all prosodic conditions. From this moment, a constant tripartite interaction starts among signal-based cues from the input (bottom-down mechanism), top-down operations like contextual expectations driven by the meaning-based cues of the input, and other top-down executive mechanisms like processing speed, inhibition and WM, which inform the processing system about prediction errors (i.e., the presence of deviants). All mechanisms together feed the attentional system, which in turn determines whether the pre-attentive involuntary attention switch will

lead to the deviant reaching full awareness, or whether the auditory processing system will continue with the processing of the following input. The successful evaluation of the deviant, and thus the activation of voluntary attention, depends on individual cognitive bandwidth. Inhibitory ability is a crucial cognitive factor to consider in understanding the observed PDR variability across individuals. Nonetheless, it is worth considering the three cognitive skills comprising cognitive profiles – inhibitory ability, processing speed, and WMC – together in an endeavour to better comprehend which cognitive operations interact with attention orienting across individuals (and how). As a general trend, correlation tests indicated that, in the current sample, individuals with better inhibitory abilities tended to be characterised by a slower processing speed and smaller WMC. Conversely, individuals with poorer inhibitory abilities tended to show a faster processing speed and larger WMC. Let us now consider the aforementioned PDR findings in the light of individual cognitive variability.

Individuals with high flanker scores, and hence strong inhibitory skills, exhibited prolonged PDRs only towards rising deviants, manifesting as a long-lasting effect, whereas they responded with quite rapid PDRs to falling and baseline neutral deviants. Despite a good ability to inhibit irrelevant information outside the current attentional focus, rising deviants broke the shield of voluntary attention, not only momentarily, but for longer time, bringing the perception of the relevant deviant to a more conscious level. Such individuals tended to exhibit a slower processing time, which gave them enough time to properly evaluate the importance of the deviant. During this time, the bottom-up mechanism interacts with the processing system and feeds it with cues coming from the signal. In the case of a rising deviant, the processing system gives it high importance, due to the deviant's high acoustic saliency and linguistic prominence. Given the established importance of the deviant, inhibition is blocked. The deviant thus enters WMC and subsequently activates voluntary attention. In the case of a falling deviant, or a deviant with baseline neutral intonation, the processing system renders the deviant's importance low, due to its (very) low acoustic saliency and linguistic prominence. Given the established low importance of the deviant, inhibition is activated, and attention is thus drawn back to the initial focus. In this latter case, the deviant's withholding from further processing is also achieved by WM resistance. Some studies on attentional control have shown that increased WM load prevents or attenuates auditory distraction by utilising a top-down control (e.g., SanMiguel et al. 2008). Here, individuals' low WMC potentially leads to high WM load, resulting in WMC resistance towards the unimportant deviant in order to minimise the disruption of storage processes (e.g., Berti & Schröger 2003, SanMiguel et al. 2008). In the case of rising deviants, such an operation is absent

potentially due to the established significance of the deviant, rendering further processing necessary.

In contrast to individuals with high flanker scores, individuals with low flanker scores, and thus weaker inhibitory skills, responded with prolonged PDRs to both rising and falling deviants, but with transient PDRs to baseline neutral deviants. This means that deviants, regardless of their prosodic marking (rising vs. falling), broke through the shield of voluntary attention. However, when a deviant did not differ prosodically from the standard numbers, owing to its low acoustic saliency and linguistic prominence, it only evoked an involuntary switch, drawing attention back to the initial focus. As the data reported in this chapter show, these individuals exhibited lower inhibitory skills and further tended to be characterised by faster processing speed and larger WMC. One potential scenario is that fast processing, in the current study, leads to inadequate evaluation of the importance of the acoustic cues characterising the deviant event. In other words, the lack of time for deviant evaluation may result in shallow processing, and thus an ineffective judgement of a deviant's importance. In this case, the deviant's inefficient evaluation, in conjunction with a weaker inhibitory mechanism and small WMC resistance,¹¹ enables the deviant event to move to later stages of orienting (regardless of prosodic marking). Another possible scenario is that these individuals, despite their fast processing speed, engage in deep processing and establish deviant importance. In this scenario, the importance of rising deviants blocks the available suppressing mechanisms and activates voluntary attention. In contrast, falling deviants, due to their attenuated auditory cues, may lead to increased listening effort. Increased listening effort, in turn, may interfere with the operations of the – already weak – suppressing mechanisms, resulting in increased (auditory) attention demands and thereby voluntary attention orienting. The disentangling of the two plausible scenarios needs further investigation; nonetheless it appears that the weaker the repressive capacity, the easier the attentional overloading.

Previous research on individual variation in auditory attention has yielded sparse and contradictory evidence. Starting with the role of WMC in auditory distractions (i.e., deviations), the present results contradict studies reporting that individuals with high WMC are less susceptible to attentional switches towards auditory distractions or deviations (for review, see Sörqvist et al. 2013, Hughes 2014). As mentioned in Sörqvist et al. (2013), the exact nature of the mechanism

¹¹Remember that these individuals have high WMC, which according to Berti & Schröger (2003) and SanMiguel et al. (2008) prevents WM load, in turn making WMC more susceptible to distractions.

that WMC taps into is still under debate, with one of the views claiming that high WMC attenuates distractions because individuals with high WMC have excellent inhibitory skills. It is important to note that in the relevant studies, a link between WMC and inhibition was presumed, but not tested directly. The present work differs from the aforementioned studies, as WMC and inhibition were measured separately, showing that high WMC does not necessarily imply high inhibition (although WM constitutes a control system; see Baddeley 2003). Instead, we have seen that individuals with high WMC are characterised by lower inhibitory skills than individuals with low WMC.

Whereas some studies investigating the role of WM in attention control have shown that higher WMC leads to fewer distraction effects, other studies have shown the opposite. The present results are compatible with the latter. More specifically, they are in line with Berti & Schröger (2003) and SanMiguel et al. (2008), who show that increased WM load (i.e., low WMC) attenuates or even prevents auditory distraction. SanMiguel et al. (2008) argue that the type of distraction influences WM load effects. Specifically, the authors mention that in the studies where WM load was found to increase distraction, distractions were actually “competing” stimuli generating task-related conflicts, as in a Stroop task. However, the distractor stimuli in the aforementioned study – or better, the deviant stimuli – were task-irrelevant, orienting attention away from the task. Likewise, in the present work, deviants were task-irrelevant, with no conflict being generated. Further, Sörqvist et al. (2013) showed in a meta-analysis that high WMC does not correlate with attenuated distraction when deviants are task-irrelevant. I am taking these findings as further evidence in support of the results reported in the present work.

To my knowledge, no previous study has directly tested the effect of inhibition or processing speed on auditory attention. The study by Keye et al. (2009) investigated the relationship among WM, inhibition, and processing speed. Nonetheless, it is important to note that Keye et al. (2009) were concerned with *visual* selective attention. Keye et al. (2009) found no support for a relation between WM and attention control or inhibition (conflict reduction). Further, the authors found a relation between high WMC and slow speed, as opposed to the present results. However, the findings of the present study are in line with the Heitz & Engle (2007) study, again on visual attention, in reporting that individuals with high WMC had faster responses than individuals with low WMC.

Considering the present findings holistically, it appears that individuals differ in the effective mechanisms they have at their disposal. The evaluation of deviants by the processing system constitutes the first crucial step in determining which further operations will be activated. Individuals may or may not have

an effective processing system, leading to successful or unsuccessful evaluation, respectively. The next step involves the activation of suppression mechanisms such as inhibition and WMC resistance. Likewise, individuals may or may not have efficient suppressing mechanisms to protect their attentional system from potential overloading. In this study, some individuals appear to have at their disposal both an efficient processing system, which evaluates the importance of the deviants successfully, and sufficient suppression mechanisms, activated when needed. When a deviant event is rendered important, suppressing operations – including inhibition and WMC resistance – are withdrawn and involuntary attention changes to voluntary attention, bringing the deviant event to conscious processing. When the event is rendered unimportant, suppressing operations are activated, blocking irrelevant distractions from later stages of orienting and protecting the attentional system from unnecessary overloading. For the remaining individuals, it is not as clear whether the processing system is efficient in evaluating the importance of deviants, but it is evident that they do not have at their disposal strong suppressing mechanisms. As a result, deviants, regardless of whether their importance has been established or not, command additional attentional resources, which maximises attentional processing load.

5.5 Summary

Pupillary data from 60 native German listeners were analysed for the present study. The findings

1. corroborate the claim that attention orienting is underpinned by an expectancy violation mechanism,
2. confirm that pupillometry is a rigorous technique for studying attention orienting,
3. index the attention orienting function of domain-final rises in speech,
4. show the relevance of temporal window of integration (TWI) length for attention orienting in language, and
5. highlight the important contribution of individual cognitive variability.

With respect to the attention orienting function of domain-final rises in German, the current results support the idea of an attentional bias towards pitch

rises, and extend the scope of this bias from a general cognitive level to the linguistic domain. In spoken language, rising pitch takes on a special role in attracting attention, also at a more conscious level, and even when this pitch rise is not attributable to a pitch accent, but rather to an edge tone (or edge tone complex). This finding strengthens the argument of Chapter 4 that the phonological status of the pitch event (head- or edge-associated) is not of primary relevance for attention orienting owing to a prevalence for holistic processing. In contrast, both the direction of the pitch contour (signal-based cue) and the appropriateness of the contour in the linguistic context (meaning-based cue) are fundamental for involuntary and voluntary attention mechanisms, respectively. In this study, rising edge tones, specifically, feed both signal- and meaning-based mechanisms, activating both pre-attentive and conscious attentional stages.

In the current data, rises and falls function in a similar way: they mark the end of a smaller sequence unit, including deviants which violate an expected pattern. Thus, rises take priority over falls precisely because of their prominence. This finding points towards an argument that the neural architecture of auditory cognition and spoken language shares a common pool of mechanisms which are crucial for the attention orienting network. This argument is revisited in Chapter 6.

As for the role of sequence length, this study shows that language benefits from a longer TWI than cognition. This is due to the complex nature of the speech signal – stimuli are longer in duration than simple sine waves. Based on the findings of this study, a TWI of at least 10 seconds is necessary for the formation of a memory trace related to regularities in standard stimulation. At the same time, a longer TWI allows the processing system to build stronger memory traces, which in turn evoke a stronger orienting response when violated.

Last but not least, the findings of this study show that the cognitive bandwidth deployed by individuals to process auditory deviances is critical both for an effective activation of voluntary attention and protection of the attentional system from potential overloading. Both mechanisms (voluntary attention and cognitive offloading) are crucial for successful speech interpretation. Importantly, individuals differ in the effective mechanisms they make use of, which affects the processing of deviances introduced in the environment. Individuals lacking such effective cognitive mechanisms may end up exhausting mental resources that might have been needed for other purposes. In spoken communication, this could entail a listener's failure to orient attention towards the most important part of the uttered message, which in turn could affect interpretation and speech planning. Individual cognitive variability is thus a crucial factor to take into consideration for processing studies, especially when the general trend in the data

represents individuals who can effectively use their attentional resources. The contribution of individual cognitive bandwidth to attention orienting is revisited in Chapter 6, in which I will bring together the findings of this and previous chapters and suggest a unified view of the cognitive and functional relevance of intonation for attention orienting in spoken language.

6 General discussion

The present work made a novel attempt to unravel the cognitive and functional relevance of intonational prominence for attention orienting in the linguistic domain. This was achieved by focusing on the temporal unfolding of the neurophysiological and pupillary underpinnings of the orienting response (OR) towards intonational rises and falls embedded in oddball sequences, along with the relevant language-specific expectations. Attention orienting has mainly been investigated within cognitive domains like vision and audition, uncovering a sensitivity of the orienting response towards the physical properties of events (visual/acoustic saliency). The novelty of the current work is that, departing from neurocognitive studies on auditory perception, it raised the question of how the function of the orienting response extends from acoustic saliency to prominence, and hence from auditory cognition to speech. In what follows, I will first summarise the main results of the two studies presented in this book (Section 6.1). Next, I will discuss the general implications of the main findings (Section 6.2), and point towards future research directions (Section 6.3). The chapter closes with concluding remarks in Section 6.4.

6.1 Summary and main observations

The aim of the study reported in Chapter 4 was to examine the neurophysiological correlates (ERPs) of attention orienting towards intonational rises and falls when presented unexpectedly in a stream of repetitive auditory stimuli. Specifically, the processing of rises and falls was investigated through the use of an oddball paradigm, where sequences of repetitive stimuli were occasionally interspersed with a deviant stimulus, allowing for elicitation of a mismatch negativity (MMN). Whereas previous oddball studies on attention towards unexpected sounds involving pitch rises were conducted on non-linguistic stimuli, this study used lexical items with natural intonation contours as stimuli. The main findings of this study can be summarised as follows:

1. The direction of the pitch contour (the signal-driven cue) is primarily relevant for involuntary attention orienting (RQ1 [rises], RQ4 [architecture]).

6 General discussion

2. Similarly, the appropriateness of the contour in the linguistic context (the context-driven cue) is primarily relevant for voluntary attention orienting (RQ3 [expectations], RQ4 [architecture]).
3. By contrast, the effect of the rise holds regardless of the phonological status of the pitch event (RQ2 [edges]). This is not predicted by standard intonational theory, where only rises on heads (i.e., pitch accents) are expected to play a role in attention orienting.

More specifically, the obtained results indicate that intonational rises play a special role in attention orienting at a pre-attentive processing stage, whereas contextual meaning (here a list of items) is essential for activating attentional resources at a conscious processing stage. This is reflected in the activation of distinct brain responses: rising intonation evokes the largest MMN; falling intonation elicits a less pronounced MMN followed by a P3 (reflecting a conscious processing stage). The data further reveal a complex interplay between the phonological status (i.e., accent/head marking vs. boundary/edge marking) and the direction of pitch change in their contribution to attention orienting. Attention is not necessarily oriented towards a specific position in prosodic structure (head or edge). Rather, it was found that the intonation contour itself, and the appropriateness of the contour in the linguistic context, are the primary cues to the two core stages of attention orienting, pre-attentive and conscious orientation respectively, whereas the status of pitch accent or edge tone plays only a minimal role. This could be either because of holistic processing, or because rises in either phonological position serve as attention orienters, in that both are meaningful rises, and not just transitions from one phonological event to another.

Chapter 5 delved into the pupillary underpinnings of attention orienting towards domain-final (the reflex of phrase final edge tones) intonational rises and falls. Using a *changing-state* oddball paradigm, in which auditory sequences of sequentially ordered (seriatim) ascending numbers (standards) were occasionally interspersed with an out-of-the-sequence number (deviant), this study investigated whether domain-final rising pitch in speech takes on a special role in attention orienting by measuring listeners' pupil dilation response (PDR). Further, this study placed a special focus on the contribution of individual cognitive variability to attention orienting. The main findings of this study can be summarised as follows:

1. Intonational rises take on a special role in attracting attention, also at a more conscious level, and even when this rise is attributable to edge tones (RQ1 [rises], RQ2 [edges], RQ3 [expectations], RQ4 [architecture]).

2. The cognitive bandwidth deployed by individuals in processing auditory deviances is critical for the effective activation of voluntary attention and protection of the attentional system from potential overloading (RQ4 [architecture]).

Elaborating on the results, intonational rises are salient, prominent, and contextually relevant (within the chosen experimental set-up), and therefore activate both involuntary and voluntary attention, which is reflected in listeners' PDRs. Given that domain-final rises and falls function in a similar way in the context of a list, namely signalling the end of a sub-sequence in a longer list, rises take priority over falls precisely because of their prominence. This finding highlights the attention orienting function of intonational rises, suggesting that domain-final rises on deviant stimuli also enhance the ability of the deviant to attract attention. Further, this finding strengthens the argument of Chapter 4 that the direction of the pitch contour and its appropriateness in the linguistic context are fundamental for attention orienting, whereas the phonological status of the pitch event is not of primary relevance. Finally, this study shows that individual cognitive variability contributes to the function of attention orienting. Cognitive variability was measured on the basis of three cognitive skills: processing speed, inhibitory ability, and working memory capacity (WMC). The results reveal that individuals differ in the effective mechanisms they tend to use, which in turn affects how they process deviances.

6.2 General implications of main findings

I now turn to the synthesis of the obtained results, offering a unified view on the cognitive and functional relevance of intonation for attention orienting, demonstrating that auditory cognition and spoken language share a common pool of mechanisms that pertain to the orienting response.

6.2.1 The relevance of intonation for the architecture of attention orienting in spoken language

Going back to the conceptual origins of OR, the hypothesis of two OR stages – involuntary and voluntary – and their attribution to different mechanisms – bottom-up and top-down – was already discussed by early philosophers, theorists, and psychologists (for more, see Chapter 2). Ivan Pavlov, the father of the OR concept, distinguished two signal systems. The first signal regulates the involuntary OR, initiated by the sensory properties of a stimulus, and the second

6 General discussion

signal system regulates the voluntary OR, evoked by complex thought processes. Importantly, the interplay between the mechanisms pertaining to OR and either low-level sensory cues or high-level cognitive operations was also attested by later psycho- and neurophysiological studies on auditory attention (for reviews, see Näätänen 1992, Näätänen et al. 2019). Essentially, the nature of the cue determines which mechanism will be fed. The different mechanisms, in turn, activate distinct attentional processing stages, resulting in involuntary or voluntary orienting.

One of the questions that the present research has raised is whether the neurocognitive architecture of attention orienting is similar across auditory cognition and spoken language (RQ4 [architecture]). The experimental findings reported in this work show that there are commonalities across the two, as the mechanisms that pertain to the different stages of the orienting response are subserved by similar neural and pupillary underpinnings.

In its most concise form, the main finding of this work is that signal-driven cues are decisive for activating an involuntary orienting response, while context-driven¹ cues are essential for activating voluntary attention orienting. In the following, I will break this finding down by jointly discussing results from the two experimental studies. Figure 6.1 presents a schematic illustration of how signal- and context-driven cues pertain to OR activation according to the present findings. Signal- and context-driven cues are represented by squares and triangles, respectively. The “square” violation represents signal-driven violations that do not fit a given contextual meaning; the “triangle” violation represents contextually meaningful signal violations that fit the surrounding context. For example, in the context of the experimental investigation reported in Chapter 4, rising deviants are linked to the square violation, while falling deviants relate to the triangle violation. In the context of the experimental investigation reported in Chapter 5, rising deviants pertain to the triangle violation as well.

The qualitatively and quantitatively distinct ERP and PDR responses show that, on the one hand, the physical properties of the sensory input (signal-driven cues) are essential for speech processing at an early pre-attentive stage, attracting involuntary attention (RQ1 [rises]). On the other hand, the meaning-based properties of the same input (context-driven cues) are crucial for speech processing at a conscious stage (RQ3 [expectations]), leading thus to voluntary attention. Both pitch rises and falls, when presented as deviants, evoked an MMN, the neural correlate that indexes pre-attentive processes related to involuntary attention

¹Context-driven and meaning-based terms have been used interchangeably throughout this book, as they refer to the semantics of a respective stimulus, as well as to meaningful aspects licensed by pragmatics and/or contextually created expectations.

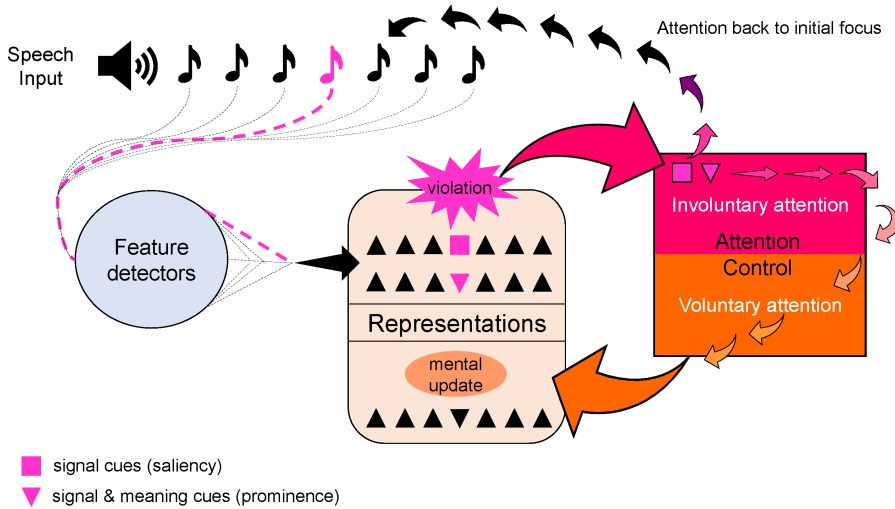


Figure 6.1: Schematic illustration of the processing routes of the OR activations, as a function of signal- or meaning-based cues. The square represents signal-driven violations; the triangle represents contextually meaningful signal violations (inspired from Näätänen et al. 2011: 12).

orienting (e.g., Näätänen et al. 2019), with intonational rises evoking the largest MMN, owing to the acoustic saliency of the signal (Chapter 4). Similarly, intonational rises and falls marking deviant numbers evoked increased PDR effects compared to deviants that do not differ intonationally from the standard numbers (Chapter 5). These results indicate that the prosodic marking of deviants (i.e., signal-driven cues) activates pre-attentive attentional resources towards the relevant violation (see Figure 6.1, where both square and triangle violations initiate an involuntary OR). The saliency level of these signal-driven cues modulates the robustness of the OR, such that the more salient the properties of the speech signal, the more robust the involuntary OR. This is in line with findings from neurophysiological and pupillometric studies on auditory perception showing that the OR is sensitive to the physical saliency of the signal, such that the more salient the physical properties of a sound, the more enhanced the OR towards it (e.g., Näätänen et al. 1978, Alain et al. 1994, Doeller et al. 2003, Rinne et al. 2005, 2006, Bach et al. 2008, Macdonald & Campbell 2011, Hsu et al. 2015, Liao et al. 2016, Wetzel et al. 2016, Marois et al. 2018).

Evidence reinforcing the fundamental importance of signal-based cues at the pre-attentive level also comes from linguistic research showing that the elicita-

6 General discussion

tion of attention-related brain responses is sensitive both to the salient acoustic contrasts in the signal and the timing of the acoustic cues (e.g., Ren et al. 2009, Tsang et al. 2011, Li & Chen 2018, Röhr et al. 2021). The present results are compatible with the aforementioned studies on different languages like Mandarin Chinese, Cantonese, and German, suggesting a cross-linguistic similarity (for more on these studies, see Chapter 3).

Furthermore, deviant falls were shown in Chapter 4 to elicit an additional P3 response, reflecting that its processing has been brought into awareness (e.g., Polich 1986). Although falls are acoustically less salient than rises, they are contextually more appropriate (within the given experimental setting), and thus contextually more prominent, when presented as deviants after repetitive rises in the context of a list. This is because this whole intonational pattern reflects one of the most natural list intonation patterns in German (for more on list intonation in German, see Chapter 3). In the experimental setting of Chapter 5, although both deviant domain-final rises and falls share a similar contextual meaning, that of signalling the end of a sub-sequence in a list, rises are linguistically more prominent. This is evidenced by the fact that deviant domain-final rises evoked more prolonged PDRs than falls, leading to voluntary attention orienting, and thus conscious processing of the relevant deviant.² These results indicate that the linguistically meaningful aspects of intonation licensed by contextually created expectations are decisive for the elicitation of voluntary OR (see illustration in Figure 6.1 where only the triangle violation that fits the surrounding context leads to a voluntary OR). This is compatible with linguistic research on the interaction between signal- and context-driven processing (e.g., Liu et al. 2016, Röhr et al. 2021) showing that voluntary attention is modulated by meaning-based expectations derived from the pragmatic context.

Therefore, just as in auditory cognition, so too in spoken language, a number of cues can activate the orienting response, with distinct cues tied to the two core attentional stages. Pre-attentively, attention is driven by the acoustic saliency of intonation, feeding bottom-up mechanisms like signal-driven expectations. Consciously, attention is driven by the meaning and the prominence relations that intonation encodes in a given context, feeding top-down mechanisms based on, e.g., language experience and context-driven expectations. Nonetheless, spoken language entails a higher level of complexity than auditory cognition, as signal-

²As discussed in Chapter 2, the latency of attention-related PDRs indicates the processing stage during which the OR is activated: transient dilations reflect pre-attentive attentional processes; more prolonged responses show that subsequent conscious processing has been activated (see Strauch et al. 2022).

driven and context-driven cues – or in other words saliency and prominence – may be combined in speakers' intonational choices.

6.2.2 The cognitive & functional contribution of intonational rises to attention orienting

The attention orienting function of rising intonation has been of particular interest for this work, due to its inherent association with prominence. As introduced in Chapter 1 (and as shown elsewhere in this book), there are many similarities between the notions of linguistic prominence and cognitive saliency. Departing from neurocognitive studies on auditory perception reporting an attentional bias towards salient rising as opposed to falling sounds, I posed the question whether this attentional bias towards rising signals extends from acoustic saliency to prominence, and hence from general cognition to spoken language (RQ1 [rises]). The present work answers this question in the positive. Concisely, the main finding is that intonational rises indeed take on a *special* role in attention orienting, particularly – but not only – at the pre-attentive stage.

That intonational rises are crucial for speech processing pre-attentively, attracting involuntary attention, is shown in the large MMN response rises evoke when presented as deviants (Chapter 4). This indicates the advantage of rises in mismatch detection, causing a greater involuntary switch of the attentional resources towards deviant rises compared to deviant falls. This holds regardless of the phonological status of the rise as a pitch accent or an edge tone. This finding highlights the pivotal role of pitch rises not only in cognition, as previous MMN studies have shown (e.g., Näätänen et al. 1978, 1980, Alain et al. 1994, Doeller et al. 2003, Hsu et al. 2015), but also in spoken language. I argue that this finding further shows that it is precisely the intrinsic acoustic saliency of rises that renders them fundamentally important for involuntary attention-related processes. In auditory perception, sounds with rising acoustic properties have been attributed a robust warning function, related to life-threatening or emergency situations, precisely because of their acoustic saliency, as they prepare the nervous system by activating attentional resources (e.g., Bach et al. 2008) which in turn elicit automated or appropriate adaptive responses (called reflexes by Sokolov 1963b, Sokolov 1963a). In speech perception, this equates to intonational rises alerting and orienting listeners' attentional resources towards, e.g., important linguistic entities. This process is crucial for effective interpretation and speech planning.

The contribution of intonational rises to attention orienting in spoken language is not only restricted to the pre-attentive stage. As we have seen, meaning-based cues are decisive for voluntary orienting, and thus for attentional processes

at a conscious stage. I showed in Chapter 4 that in speech, the acoustic saliency of intonational rises (and thus their attention orienting function) is mitigated by contextual meaning: when intonational rises are contextually inappropriate, they do not attract attention at a conscious level. By contrast, falls, which are not salient but contextually relevant, and thus prominent, do attract voluntary attention. Further, Chapter 5 showed that when both intonational rises and falls are contextually relevant, rises trump falls. I argue that this finding highlights the attention orienting function of intonational rises, even when they are attributable to edge tones, suggesting that when rises and falls function similarly in a given context, rises on deviant stimuli enhance the ability of these deviants to attract attention also at a more conscious level. Rises in this case are not only acoustically more salient, but also linguistically more prominent than falls. This is in line with studies on the neural processing of linguistically meaningful pitch variations in different languages (e.g., Tsang et al. 2011, Ventura et al. 2020, Röhr et al. 2021). For example, Tsang et al. (2011) showed in a study on the pre-attentive processing of Cantonese tones, that MMN is sensitive to pitch contour direction and height. Ventura et al. (2020) found in an EEG study that rising pitch accents guide attention towards semantic incongruencies, as they lead to more elaborate processing of the focused information. Likewise, Röhr et al. (2021), investigating the signal- and context-driven effects of German prosodic prominence in an EEG study, reported a P3 elicitation, a neural correlate of conscious attention, to intonational rises on the stressed syllable of the critical words, and claimed that it is the linguistically “*meaningful* rises that orient attention, rather than rises *per se*” (2021: 44).

Another question I raised in this work is whether rises at the edges of constituents attract attention to the same extent as accentual rises do (RQ2 [edges]). The present findings also answer this question in the positive. The experimental investigations reported in Chapters 4 and 5 both provide evidence for the attention orienting function of intonational rises, even when they are attributable to edge tones. That both accentual and boundary intonational rises have been found to function as attention orienting devices in the present research challenges the typological prediction posited by the autosegmental-metrical (AM) theory of intonational phonology (e.g., Ladd 2008, Arvaniti 2022, Grice 2022a), where pitch accents are strictly associated with a prominence-cueing function, and edge tones are associated with a mere phrasing function. Serial recall studies (Savino et al. 2020, Röhr, Savino, et al. 2022, Grice et al. 2024) have already provided first indications against this strict dichotomy, in reporting that rising edge tones marking the final item of non-final triplets boost recall accuracy of the whole triplet, orienting attention to the whole domain. The two present ex-

perimental investigations corroborate these findings by showing that not only accentual but also edge tone-related intonational rises can cue prominence, by orienting attention towards a deviant bearing them. All these findings posit a challenge for important aspects of prosodic theory and typology in pointing towards a prominence-cueing function of rising edge tones.

Overall, the current results contribute to our understanding of how intonation affects listeners' detection of prediction-violations or mismatches, and thus the orienting of attention towards them by showing i) that intonational rises indeed make information more salient, attracting attention pre-attentively (serving like warning cues), and ii) that the attention orienting function of rises is mediated both by prominence and the expectations within a given context, leading to voluntary attention orienting.

6.2.3 The role of expectations in attention orienting

It appears that the auditory and linguistic processing systems share common features relating to attention orienting, with expectations (i.e., predictions) being a driving force. A question that this work has put forward in relation to expectations is whether language- and/or context-specific expectations affect attention orienting (RQ2 [expectations]). As Chapter 2 thoroughly reviewed, one of the mechanisms deemed to underlie attention orienting is the *expectancy violation* mechanism, suggesting that the precondition for an OR is the violation of rule-based expectancies (e.g., Näätänen 1990, Sussman & Winkler 2001, Vachon et al. 2012).

According to Näätänen et al.'s (2011) prevalent model, the auditory processing system develops sensory-memory representations related to the regularities between successive sound events in the environment (see the features of the auditory input integrated in the formation of representations in Figure 6.1). On the basis of these representations, the processing system forms expectations by predicting the properties of upcoming sound events. When these anticipatory predictions are violated because the properties of the upcoming event deviate from what is expected, attention-calling mechanisms are activated. As we already saw, the first attention-calling stage is the involuntary attention switch, activated by signal-driven cues. After the involuntary OR towards the auditory deviance, this deviance can be brought into awareness, either by a strong attention-calling process or through the formation of an attentional trace. This trace depends on voluntary maintenance by the control mechanisms (voluntary OR). When it is voluntarily maintained, it is available to top-down processes as well as to long-term memory, leading to conscious perception and semantic activation (see Sec-

tion 2.3). The findings in the present work confirm that the same expectancy violation mechanism activates the subsequent attention-calling mechanisms in spoken language.

Let me exemplify this last point by referring to the data from Chapter 4. Here, listeners were presented with classic oddball lists, namely sequences of repetitive sounds (standards) occasionally interspersed with a rare sound (deviant). Specifically, rising and falling intonation alternated as standard/deviant sounds across different conditions. Rising and falling intonation was introduced on real lexical items in German referring to food items, giving rise to a list context and the relevant language-specific expectations. In this study, two layers of expectations were formed. First, our auditory processing system is able to predict prospective sounds by detecting regularities in the input. Therefore, listeners, after being exposed to the repetitive sequence of stimuli, either with rising or falling pitch, predict that the next event would bear the same pitch contour (signal-driven expectations). When this was not the case, an MMN was elicited, indicating an involuntary attention switch towards the deviant. Remember that all deviants elicited an MMN – this means that all deviants, regardless of their acoustic saliency, were able to activate the first attention-calling stage,³ simply by violating signal-driven expectations. Predictions can, however, also arise from the linguistic interpretation of the context (context-driven expectations): the oddball paradigm resembles a list, and, additionally, the natural pitch contours used on real words enhanced the list context. Thus, when rising stimulation was presented repetitively (standard stimuli), as a natural and appropriate intonation on non-final list items, this appeared to elicit additional contextual predictions, such as the anticipation of the end of the list. Therefore, when the falling deviant was presented, it validated the context-driven expectations, as it is a cue to finality – even though it violated the signal-driven expectations. This is reflected in the MMN-P3 complex elicited by this deviant. Pre-attentively, the falling deviant, violating the signal-driven expectations, activates an involuntary OR, shown in MMN, but subsequently, manages to be voluntarily maintained and move into a conscious processing stage by validating the context-driven expectations, shown in the P3. These context-driven expectations were not formed when falling pitch was repetitively presented and occasionally interrupted by rising deviants, probably because the repetitive standard falling stimulation is already less expected on non-final items in the context of a list, blocking the generation of predictions over and above the pure signal.

³However, the more salient the intonational contour, the more robust the MMN response.

Let us now move to the data from Chapter 5. In this study, listeners were presented with seriatim (i.e., sequentially ordered) ascending numeric lists (standards), occasionally interspersed with an out-of-the-sequence number (deviant). This deviant number featured either the same intonation as the standard numbers (neutral), or either final rising or final falling intonation. The pupillometric data show that all deviant numbers did elicit a PDR, regardless of intonation. This finding further corroborates the claim that attention orienting is underpinned by an expectancy violation mechanism. The numeric lists inherently give rise to expectations about forthcoming numbers. The processing system forms memory representations by detecting the regularity in the environment, that is, the seriatim pattern, and predicts that the next number will follow the same pattern. This pattern was violated, once the deviant number was presented, causing an OR as reflected in the PDRs. Listeners were not only exposed to the seriatim pattern, but they were also exposed to the signal-driven cues with which this pattern was produced. Therefore, listeners not only anticipated a following stepwise ascending number, but they also expected this number to feature neutral intonation, given that all previous numbers were produced with this realisation. When the deviant violated both the seriatim and the intonational pattern, the OR response was stronger, as reflected in the increased PDRs. This result shows that signal-driven cues garner more robust attentional resources towards the violation. Besides the seriatim and the signal-driven expectations, the listeners appeared to have formed a third layer of expectation, which arose from the linguistic meaning of intonation (context-driven expectations). More specifically, the neutral intonation of the standard stimulation is a natural list intonational pattern for non-final items in German. Thus, the expectations regarding a forthcoming number were enhanced by the linguistic meaning of the neutral intonation, indicating that the list was not over yet. Numeric lists are not infinite, however, and listeners therefore also expected that the list would end at some point, and would have listened out for the corresponding intonational cue. Now, both final rises and falls can mark the end of a smaller or larger unit. Thus, when final rises and falls were presented, they, on the one hand, validated the anticipation of the end of the list. On the other hand, the deviant was included in the same unit as the standard numbers, thereby violating the anticipated seriatim and signal-driven patterns. The partial validation of the context-driven expectation, along with the violations of the other two expectation layers, led to a conscious OR.

These findings are compatible with previous linguistic research which has shown that the processing or the perception of prosodic mismatches is shaped by expectations licensed both by signal properties and meaningful aspects of prosody within a pragmatic context (e.g., Liu et al. 2016, Röhr et al. 2020, 2021,

Röhr, Baumann & Grice 2022). Studies have also shown that the processing of such mismatches can be affected by language-specific expectations (e.g., Chandrasekaran et al. 2009, Ventura et al. 2020, Grice et al. 2025). The present results are also in line with studies reporting that prior linguistic- or discourse-based expectations, or even recent speech experience, can overwrite signal-driven expectations at later processing stages (e.g., Bishop 2013, Kakouros et al. 2018, Roettger & Rimland 2020). Röhr et al. (2021) have claimed that attention orienting related to predictive processing is reflected in discrete stages in the construction of a mental representation during real-time comprehension, based on the fact that the amount of attention drawn to a stimulus is defined by signal-based cues combined with context-driven expectations.

The present findings contribute to the argument that the OR in speech is indeed underpinned by the expectancy violation mechanism proposed for the auditory domain. This is in accord with what has been suggested about the general functions of the linguistic processing system by Schumacher et al. (2015). As reviewed in Chapter 3, Schumacher et al. (2015) claim that the linguistic processing system, like the auditory processing system, constantly develops mental representations, related to entities in the linguistic environment. Based on these representations, predictions for the upcoming entities are derived. Prediction errors (Friston 2010) in language, which occur from item-context mismatches, have further been found to elicit detection and attention-calling processes (for more details on the architecture of the prediction-updating cycle, see also Röhr et al. 2021, Repp & Schumacher 2023). The more enhanced the violation of a prediction, the higher the processing cost for an entity (or, the more difficult its reconciliation with prior discourse; see Röhr et al. 2021). After the detection of violations, the mental model needs to update its representations. More specifically, as reported in Röhr et al. (2021), if listeners allocated their attention towards some prosodic mismatch, topic shift, or novel information after detecting it, they need to incorporate the new knowledge in their representations by repairing, reorganising, and updating their mental model (see also, Toepel et al. 2007, Dimitrova et al. 2012, Brouwer & Hoeks 2013, Schumacher et al. 2015).

6.2.4 **Neural and pupillary underpinnings of attention orienting**

Let us now zoom in further on evidence for commonalities in the architecture of auditory cognition and spoken language with regard to attention orienting (RQ4 [architecture]). More specifically, I will discuss the neural and pupillary underpinnings of attention orienting across the two domains. In general, the literature review in Chapter 2 elucidated that OR is expressed in multiple physiological,

psychological, motoric, and neural exponents. It has emerged that the MMN and P3 ERP components are rigorous correlates of attention orienting in auditory perception, reflecting the route from involuntary to voluntary orienting (e.g., Escera et al. 1998, Näätänen 2001, Hsu et al. 2023). Similarly, PDR was shown to be a valid index of auditory attention orienting, highlighting a striking resemblance between MMN/P3 components and PDR composites (e.g., Nieuwenhuis et al. 2011, Wang & Munoz 2015, Johansson & Balkenius 2018, Marois et al. 2018, Alamia et al. 2019, Strauch et al. 2022, Contier et al. 2024).

The present work has indicated that attention orienting in spoken language is manifested in the same neural and pupillary exponents. This is a crucial finding for two reasons. First, it reinforces the credibility of these OR correlates by attesting their presence in a different domain. Second, it shows that the function, or the architecture, of the orienting response is subserved by the same neural pathways and pupillary underpinnings across domains.

Overall, the common pool of mechanisms crucial for the orienting response in both the auditory realm and in speech is reflected in a number of ways, such as

1. the involuntary & voluntary OR stages, which are initiated by a similar interplay between distinct mechanisms (bottom-up & top-down) and different cues (signal-driven & context-driven);
2. the expectancy violation mechanism, which is responsible for activating attention-calling mechanisms like OR; and
3. the neural and pupillary underpinnings that subserve the orienting network.

6.2.5 Individual variability and other top-down operations

A final point I will discuss is the role of individual variability in the orienting response. Chapter 5 explored the impact of listener-specific cognitive variability on attention orienting. The aim was to better understand how other cognitive operations interact with attention orienting across individuals. In its most concise form, the main finding is that individuals do vary in the use or control of their attentional resources towards incoming information, utilising different top-down operations to different extents. Figure 6.2 presents a schematic illustration of other cognitive operations involved in the process of involuntary attention transitioning to voluntary orienting as a function of individual variability.

6 General discussion

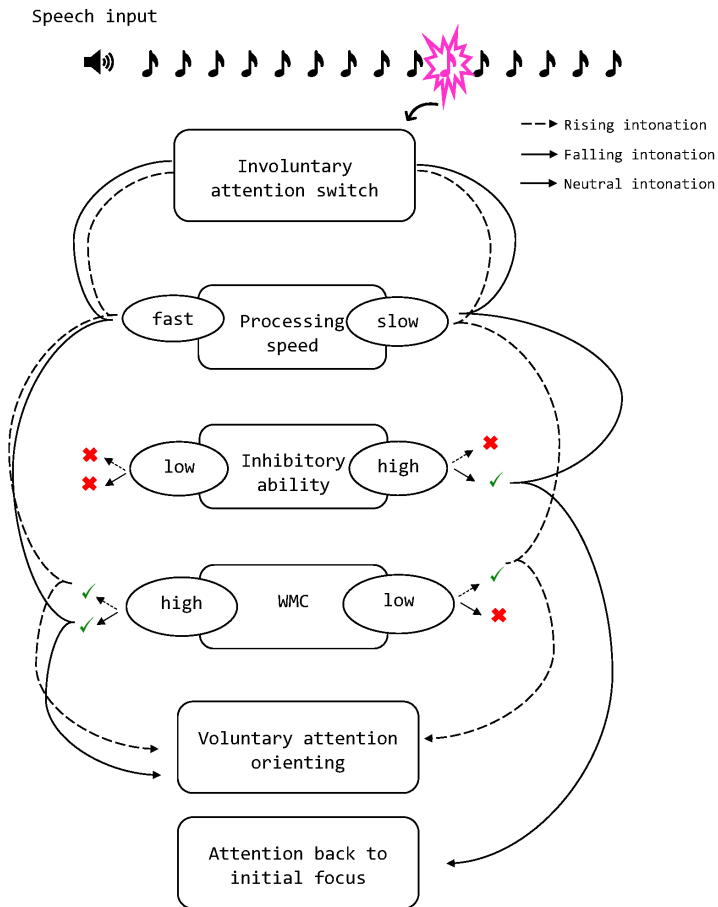


Figure 6.2: Schematic illustration of other top-down operations pertaining to attention orienting. The figure presents sets of possible activations of other cognitive operations leading from involuntary to voluntary orienting, as a function of individual cognitive profiles.

That the cognitive bandwidth deployed by individuals in processing auditory deviances in spoken language is critical for an effective activation of voluntary attention (and protection of the attentional system from potential overloading) is reflected in the observed PDR variation across individual cognitive profiles. In particular, Chapter 5 reports that all deviants elicited PDR responses. This indicates that, when deviants were presented, listener attention was involuntarily oriented towards them, regardless of intonation. Nonetheless, the three into-

national patterns marking the deviants evoked different PDR modulations as a function of individual cognitive variability.

The participants in this study tended to form two subgroups with regard to cognitive variability, operationalised with measures of inhibitory ability, processing time, and WMC. The first group of individuals (see paths represented by arrows on the right side of Figure 6.2) appeared to have both an efficient processing system, which evaluates the importance of deviants successfully, and sufficient suppressing mechanisms, which were activated when needed. That is, when a deviant event, e.g., an intonational rise, is salient and/or prominent, the processing system renders it important. Subsequently, suppressing operations, including inhibition and WMC resistance, are withdrawn and involuntary attention gives way to voluntary attention, bringing the deviant in conscious processing. When the saliency and/or prominence status of the event is low, the processing system renders it unimportant. Consequently, suppressing operations are activated, blocking irrelevant distractions from later stages of orienting and thereby protecting the attentional system from unnecessary overloading.

For the second group of individuals (see paths represented by arrows on the left side of Figure 6.2), it is not as clear whether their processing system is efficient in evaluating the importance of deviants. Nevertheless, it is evident that these individuals do not have strong suppressing mechanisms. As a result, deviants garner additional conscious attentional resources regardless of whether their importance has been established or not, thereby maximising attentional processing load.

On the basis of these results I propose that a deviant's presentation causes, at first, an involuntary attention switch across all prosodic conditions. From this moment onwards, other top-down operations, like processing speed, inhibitory ability, and WMC, cooperate and continuously interact with the attentional system. The attentional system, in turn, determines whether the pre-attentive involuntary attention switch will lead to full awareness of the deviant, or whether the processing system will continue with the processing of the following input. As we saw, the successful evaluation of the deviant, and thus the activation of voluntary attention, depends on the individual's cognitive bandwidth. A deviant's evaluation by the processing system constitutes the first crucial step in determining which further operations will follow. Individuals may or may not have an effective processing system, leading to either successful or unsuccessful evaluation. The next step involves the activation of suppressing mechanisms like inhibition and WMC resistance. Inhibitory abilities underlie what are perhaps the most crucial operations in protecting the attentional system from potential

6 *General discussion*

overloading. Likewise, as we saw, individuals may or may not have efficient suppressing mechanisms to protect their attentional system. The combination of an effective processing system and sufficient suppressing mechanisms is essential, as failure in responding to an important sound change might be life-threatening, but responding to every sound change might exhaust mental resources that are needed elsewhere. In spoken communication this could entail a listener's failure to orient attention towards the most important part of a message, which in turn can affect successful interpretation and speech planning.

As mentioned in Chapter 5, research on individual variability in auditory attention is scattered, and, to my knowledge, no previous study has directly tested the effect of inhibition or processing speed on auditory attention. With regard to WMC, the present results are compatible with Berti & Schröger (2003) and SanMiguel et al. (2008) in reporting that increased working memory (WM) load, and thus low WMC, attenuates or even prevents auditory distraction. Nonetheless, the present findings stand in contrast with other studies reporting that listeners with high WMC are less susceptible to auditory deviances (for review, see Sörqvist et al. 2013, Hughes 2014). SanMiguel et al. (2008) claim that WM load effects are influenced by the type of the auditory deviant or distraction. In this respect, SanMiguel et al. (2008) argue that, in studies where WM load (i.e., low WMC) was found to increase attention towards auditory deviants, those deviants actually were “competing” stimuli generating task-related conflicts, as in a Stroop task. However, in SanMiguel et al. (2008), the deviant stimuli were task-irrelevant, orienting attention away from the task. In a similar vein, Sörqvist et al. (2013), showed in a meta-analysis that high WMC and attenuated distraction do not correlate when deviants are task-irrelevant. Given that deviants were also task-irrelevant in the present work, with no conflict being generated, I take these findings as further evidence in support of the present results. Finally, with regard to processing speed and WMC correlation, our results are in line with a study on visual attention by Heitz & Engle (2007) reporting that individuals with high WMC had faster responses than individuals with low WMC.

Overall, the present findings on PDR contribute to research on individual variability by exploring three different cognitive skills: inhibition, processing speed, and WMC. The findings highlight the importance of taking individual cognitive variability into consideration in processing studies, and offer a starting point for further work through the cognitive test battery used for assessing the aforementioned cognitive operations.

6.3 Limitations and directions for future work

The present work provided empirical findings that contribute to a better understanding of the relevance of intonation for attention orienting in spoken language. Thus, the current results form a solid basis for future research. Nevertheless, like many other studies, the current experimental designs have some limitations. To this end, in what follows, I propose directions for future work that could deepen our understanding of attention orienting in spoken language and the impact of intonational prominence on it.

The present work investigated the temporal unfolding of the neural and pupillary underpinnings of attention orienting towards intonational rises and falls embedded in oddball sequences. The construction of the oddball sequences in the two present studies included real words, food items in Chapter 4 and numbers in Chapter 5, simulating a more natural speech context (that of a list) compared to previous neurocognitive studies. Nonetheless, this context was still very much controlled. To this end, I propose that future research should focus on the investigation of the link between intonation and attention orienting in more ecologically valid contexts. One idea could be the use of semi-spontaneous memory games. For instance, listeners could be presented auditorily with people playing a game like *I packed my bag* (Brandreth 1986). In the most common version of this game, a speaker starts by saying “I packed my bag and in it I put...” and refers to an object. The next speaker repeats the sentence with the original object and adds one more. The very next speaker repeats the sentence with the original object, the object named by the second speaker, and adds another one, and so forth. The structure of this game offers an excellent test case, as it creates memory traces of patterns which, of course, can be violated by introducing deviant items in the narrative, which, in turn, can be manipulated prosodically. Scenarios other than packing a bag can, of course, also be used, such as “I went to the zoo and saw...” (Wallace & Kauffman 1973), “I went on a picnic and brought...” (Dynes 2017), “I went to the shops and bought...” (Lauchlan & Carrigan 2013), or any other context. A more radical idea would be to test even more naturalistic settings, in presenting listeners with a podcast topic or an audio book and manipulate the intonation of some utterances in expected (congruent) and unexpected (incongruent) ways. The unexpected manipulations could serve as deviants, investigating whether or how attention is oriented towards such manipulations compared to the expected manipulations which could serve as standards.

Considering a different type of limitation, this work unravelled the cognitive and functional contribution of intonational rises, attributable both to pitch accents and edge tones, to attention orienting in German, a West Germanic lan-

guage. In order to test the biological basis of the attention orienting function of rises, as well as the role of edge tones in the orienting response, non-West Germanic languages also need to be investigated. Maltese and/or Maltese English (MalTE) offer two excellent test cases for future research. Due to historical and demographic shifts, Maltese and MalTE have been in contact for many years, and they co-exist in the repertoire of most Maltese speakers (Vella 2013). What makes Maltese and/or MalTE particularly interesting is their prosodic characteristics. First, both Maltese and MalTE are reported to have weight-sensitive lexical stress (e.g., Vella 1995, 2009). Most importantly, although both languages are described as head prominence languages (Jun 2014a), meaning that they feature regular pitch accents to cue prominence, they also exhibit frequent H(igh) tones on initial unstressed syllables (e.g., Vella 2007, 2011). Thus, unlike German where lexical (stress) and postlexical (intonation) prominence co-occur, the two levels of prominence do not appear to be straightforwardly linked in Maltese and MalTE (Lialiou et al. 2023). This enables words in Maltese and MalTE to feature more than one prominence simultaneously: lexical stress prominence as well as intonational prominence on the initial syllable. This potentially raises questions about the perceptual status of prominence in Maltese and MalTE, as lexical stress and intonational prominence might serve as competing cues (Lialiou et al. 2023). Therefore, the investigation of the attention orienting function of initial rises in Maltese and MalTE, and its comparison to the attention orienting function of regular accentual rises, could shed light not only on the general importance of rises in attention, but more importantly on the functional contribution of rises at the edges of constituents to prominence.

Another interesting test case could be Greek. Typologically, Greek, too, is a head prominence language (e.g., Jun 2014a). Nonetheless, recent results on the perception of prominence in Greek show that L(ow) accents are equally likely to be judged as prominent as other rising accents when they appear in a nuclear position (Orrico et al. 2024). Orrico et al. (2024) suggest that their results indicate that it is the strong metrical position that drives the prominence perception in Greek, rather than the mere acoustic properties of the intonational events. It has been shown in Chapter 4 that listener attention is not always attracted towards the most acoustically salient entity. Although German is a language where prominence is very often encoded in the form of intonational rises, listener attention was voluntarily allocated to deviant falls. This voluntary orienting towards deviant falls was not driven by acoustic saliency, but rather by linguistic prominence and pragmatic criteria, in the form of language- and context-driven expectations. Orrico et al. (2024) mention that their results on Greek show precisely that Greek listeners' prominence judgments are not necessarily driven by

acoustic saliency, but rather by higher order phonological and pragmatic factors. Therefore, testing the relevance of intonation for the orienting response in Greek could potentially shed more light on whether the biological importance of rises as an attentional cue still holds in a language that may not prioritise the acoustic signal to cue prominence, and could thereby elucidate the role of language-specificity in interpreting cues and in attention orienting.

A final point that deserves further investigation is the role of individual variability in attention orienting. As already mentioned in the previous section, Chapter 5 made a novel attempt to unravel the importance of individual variability in attention orienting with regard to other cognitive operations such as processing speed, inhibitory ability, and WMC. Nonetheless, other cognitive operations may also be relevant for a better understanding of the orienting response across different listeners. In this regard, a further shortcoming of the current work is that Chapter 4 did not look at all into individual variability. Given the promising findings of Chapter 5, I propose that future research in the direction of individual variability is essential, especially when group level results are driven by a specific subgroup of the registered sample. The present work provides a basis for future research by offering a cognitive test battery that includes assessments for the aforementioned cognitive operations. Certainly, future work further testing the relevance of these cognitive operations is warranted. Given that the bigger picture on individual variability remains blurry, researchers who endeavour to better understand individual processing variability should consider investigating further cognitive operations and other skills, such as pragmatic skills.

6.4 Conclusion

To conclude, the present work made a novel attempt to unravel the relevance of intonation for attention orienting in the linguistic domain. It provided experimental data using ERPs and pupillometry, which were utilised to examine the route from involuntary to voluntary attention. Taken together, the findings from the experiments reported in this book indicate that both saliency and prominence, encoded through intonational choices, impact the orienting response by modulating different stages of attention. The current work increases our knowledge in relation to the orienting response in the linguistic domain, and takes us one step closer to a full understanding of how linguistic complexity, with its interactions between signal- and context-driven cues, pertains to the attentional network. Lastly, the present research corroborates the idea of an attentional bias towards intonational rises, associated with both heads and edges, and extends it, for the first time, from acoustic saliency to prominence, and hence from general cognition to spoken language.

Appendix A: Pilot study – Button press

The aim of this pilot study was to test the design of the stimuli used in the pupillometric experiment reported in Chapter 5, ensuring that the introduced deviances in the numeric sequences were indeed perceivable. For the pilot study, an explicit deviant-detection task was employed using a button press paradigm. Participants in the pilot study also performed the cognitive test battery measuring individual variability, so that the respective implementations could be tested as well. A different sample of listeners was necessarily recruited for the pilot and the pupillometry studies, given that the listeners who took part in the pilot study were aware of the presence of deviants in the items.

A.1 Participants

Thirty native speakers of German (19 female, 11 male), aged between 20 and 34 years (mean age = 24.7 years, $SD = 3.9$), took part in the study. Participants provided written informed consent in accordance with the Declaration of Helsinki and in compliance with the ethics clearance from the Ethics Board of the *Deutsche Gesellschaft für Sprachwissenschaft* (DGfS). Participants received reimbursement for their participation (either course credit or monetary compensation). None of them reported any speech, hearing, or neurological impairments.

A.2 Speech materials

The design and the acoustic characterisation of the speech materials tested in this pilot study are those used in the pupillometry study reported in Section 5.2.2 and Section 5.2.3.

A.3 Experimental procedures

A.3.1 Button press

Each experimental session consisted of the pilot task, i.e., a button press task, succeeded by a battery of cognitive tests, always in the following order: a version of

A Pilot study – Button press

the flanker task (measuring inhibitory ability), a version of the odd-man-out task (measuring processing speed), and a version of the digit span task (measuring WMC; see Section 5.2.4 for more details on these tasks).¹ After a briefing about the structure of the experimental session, participants started with the deviant-detection task in a button press paradigm. Participants were seated in a quiet lab room in front of a computer monitor and a keyboard, wearing headphones. The experimenter informed the participants that they would be presented auditorily with numeric sequences. Each experimental trial started with a screen presenting a black fixation cross for 1500 ms. Participants were instructed to fixate the black cross that would appear on the screen, and were informed that as soon as the cross turned green, they would hear a sequence. The green cross remained on the screen throughout the audio file’s duration. Participants were also informed that some of the sequences would include some out-of-the-sequence numbers, and were asked to press the SPACE button on the keyboard as fast as possible when they heard the deviant number. A time-out was set at the end of the audio file. Each trial ended with a blank screen for 1000 ms. Written instructions were also provided. The task started with a practice phase of five items. After the practice phase, participants had the option to ask potential clarification questions. The task consisted of three blocks of 24 items each. The task was approximately 35 minutes long. Figure A.1 illustrates the structure of an experimental trial.

A.3.1.1 Cognitive test battery

For the description of each task in the battery test, see Section 5.2.4.

A.4 Data processing and statistical analyses

Data processing and statistical analyses were conducted in R, version 4.1.2 (R Core Team 2023), using the R packages *brms* 2.17.0 (Bürkner et al. 2023), *ggplot2* 3.3.5 (Wickham 2016), and *tidyverse* 1.3.1 (Wickham et al. 2019). For reproducibility, data and scripts have been made available at <https://osf.io/nhy4c>.

¹The button press, the flanker, and the odd-man-out tasks were implemented in *OpenSesame* (Mathôt et al. 2012). The digit span task was implemented in *SoSci Survey* (Leiner 2024). In all these tasks, participants could take an optional short break between the blocks, and during the practice phase they received immediate feedback on the screen.

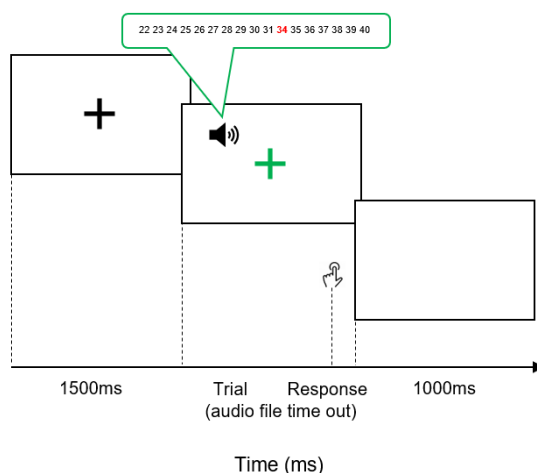


Figure A.1: Schematic illustration of an experimental trial in the button press task.

A.4.1 Data pre-processing

A.4.1.1 Button press task

In the pilot task, accuracy (correct/incorrect responses coded as 1/0) as well as response times (in ms) were recorded for each trial, and participants' mean accuracy and mean correct response time were computed. Every trial in which participants pressed the response button after the onset of the deviant in trials introducing arithmetic deviances (the experimental items) was considered a correct response. Simultaneously, the lack of a button press in trials which did not include deviances, i.e., the filler items, were considered correct responses. Every trial where participants pressed the response button before the presentation of the deviant in the sequence, did not press the response button at all, although the sequence included a deviant, or pressed the response button although the sequence did not include a deviant were taken to be incorrect responses.² Correct response time was computed as the time from the onset of the deviant number until the press of the response button on every trial including a deviant.

A.4.1.2 Cognitive test battery

For the pre-processing of the cognitive scores, see Section 5.2.5.

²One participant was excluded from the analyses, as their overall accuracy in the filler items was 14%.

A.4.2 Inference criteria

The statistical evaluation of the pilot results is divided into *confirmatory* and *exploratory* analyses. In the confirmatory part, I tested i) whether deviant numbers are indeed detected, and ii) whether rising intonation, due to its reported attention orienting function, facilitates deviant recognition, and thus task performance, compared to falling or neutral intonation.

In the exploratory part, I examined two-way interactions between the main variable, that is prosodic condition, and two other predictors. Specifically, given the two different item lengths (medium and long), which controlled for the deviant position, I explored whether sequence length affects accuracy and correct response times, and whether length interacts with the prosodic conditions. Further, I investigated if and how individual cognitive variability affects the processing of the introduced deviances produced with different intonational patterns. Finally, I investigated whether there is a speed–accuracy trade-off in the participants' performance.

A.4.2.1 Confirmatory analyses

Correct response times (in ms) and accuracy (correct/incorrect response, coded as 1/0) were modelled by fitting separate Bayesian hierarchical regression models (linear hierarchical regression for response times, and Bernoulli logistic regression for accuracy) as a function of PROSODIC CONDITION. Treatment contrast was used to code the predictor PROSODIC CONDITION (levels: neutral/rise/fall) with the level neutral serving as the reference. Random effects for both SUBJECTS and ITEMS included full variance–covariance matrices (e.g., Barr et al. 2013).

A.4.2.2 Exploratory analyses

A.4.2.2.1 Sequence length

Correct response time (in ms) and accuracy (correct/incorrect response) were modelled separately as a function of two-way interactions between PROSODIC CONDITION and LENGTH. Similarly to PROSODIC CONDITION, treatment contrast was used to code the predictor LENGTH (levels: long/medium), with the long stimuli serving as the reference level. Random effects for both SUBJECTS and ITEMS included full variance–covariance matrices (e.g., Barr et al. 2013).

A.4.2.2.2 Cognitive variability

The continuous cognitive scores resulting from the three cognitive tests were z-scored before entering any statistical models. Subsequently, they were tested for correlations in order to find out whether participants' skills are interrelated. In the case of a correlation among the three scores, only one of the three scores would be selected for the subsequent analyses. Correct response time (in ms) and accuracy (correct/incorrect response) were modelled separately as a function of two-way interactions between PROSODIC CONDITION and cognitive score(s). Random effects for both SUBJECTS and ITEMS included full variance–covariance matrices (e.g., Barr et al. 2013).

A.4.2.2.3 Speed–accuracy trade-off

Finally, the speed–accuracy trade-off was also examined by modelling accuracy (correct/incorrect response) as a function of a two-way interaction between PROSODIC CONDITION and SPEED (i.e., response time, z-scored). Here, response times of both correct and incorrect responses were included. Random effects for both SUBJECTS and ITEMS included full variance–covariance matrices (e.g., Barr et al. 2013).

All models used weakly informative priors for all parameters (the full prior specification can be found in the script provided on OSF), allowing for a wide range of effect sizes while controlling for unreasonably large effects. All models ran with four chains and 4000 iterations, with a warm-up period of 2000 iterations, and they all converged: there were no divergent transitions and all \hat{R} s were close to 1, showing that chains mixed without issues. Model fits were also visually inspected using the posterior predictive check function. In the following section, inferences are drawn using the posterior distributions of the parameters. For this, the following are reported: posterior estimates, standard errors (SE), the low and high boundaries of the 90% credible interval (CrI) of the estimate, and the posterior probability that the estimate falls on one side of zero (e.g., $P(\beta < 0) = 0.95$). When almost all of the posterior mass for an estimate lies on one side of zero, zero is not included in the 90% CrI (by a reasonably clear margin), and the posterior probability P is close to one, the effect is considered reliable.

A.5 Results

Figure A.2 illustrates accuracy on deviant identification (left panel), and response times of the accurate responses (right panel) across the three prosodic conditions

A Pilot study – Button press

(neutral, rise, fall). Violin plots visualise distributions of values as kernel density plots. On top of the violins, mean accuracy and mean response times are plotted using dot plots. Dots in different colours illustrate mean scores of each individual participant per prosodic condition, while the large triangles depict mean values of each distribution at the group level. Visual inspection of the data shows that the overall accuracy is at ceiling across all prosodic conditions (neutral = 94%, rise = 97%, fall = 96%), and that mean response times do not differ considerably across prosodic conditions (neutral = 919 ms, rise = 942 ms, fall = 952 ms).

In the following, I first report on the confirmatory analyses, and subsequently explore interactions between the main variable and sequence length as well as individual cognitive variability. Finally, I present an analysis of potential speed–accuracy trade-offs in individual performance.

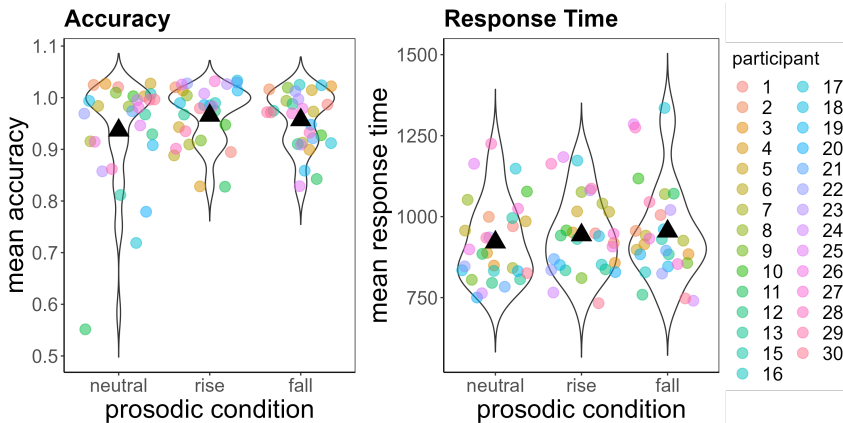


Figure A.2: Distribution of accuracy on deviant identification in the numeric sequences (left panel) and mean response times of accurate responses (right panel) across the three prosodic conditions (neutral, rise, fall).

A.5.1 Confirmatory analyses

Model outputs suggest that prosodic condition did not reliably affect either accuracy or correct response times, as they did not provide evidence for reliable differences among conditions. Estimated contrast details are given in Table A.1.

Table A.1: Tabular overview of modelling accuracy (a) and response time (b) as a function of prosodic condition.

(a) Accuracy						
comparison	β	SE	Low CrI	High CrI	Evid. Ratio	Post. Prob
neutral vs. rise	0.56	0.43	-0.15	1.26	9.81	$P(\beta > 0) = 0.91$
neutral vs. fall	0.41	0.40	-0.23	1.08	5.38	$P(\beta > 0) = 0.84$
rise vs. fall	0.15	0.49	-0.68	0.96	1.71	$P(\beta > 0) = 0.63$

(b) Response time						
comparison	β	SE	Low CrI	High CrI	Evid. Ratio	Post. Prob
neutral vs. rise	0.02	0.02	-0.02	0.06	3.96	$P(\beta > 0) = 0.80$
neutral vs. fall	0.03	0.02	-0.01	0.07	7.96	$P(\beta > 0) = 0.89$
rise vs. fall	-0.01	0.03	-0.05	0.03	0.61	$P(\beta > 0) = 0.38$

A.5.2 Exploratory analyses

A.5.2.1 Sequence length

First, it was tested whether sequence length had an effect on both dependent variables (i.e., accuracy and correct response times) as well as whether there was an interaction with prosodic condition. Models on both accuracy and correct response times indicate that there is no reliable evidence for either an effect of length or an interaction between prosodic condition and sequence length. Estimated comparison details for accuracy are presented in Table A.2, and for correct response times in Table A.3.

A.5.2.2 Cognitive variability

Subsequently, it was explored whether individual cognitive variability interacts with the prosodic structure of deviants, affecting task performance. Before moving on to the modelling results, I will report on the cognitive profile of the individuals. Table A.4 presents raw mean scores per individual participant across the three cognitive tasks.

Higher scores in the flanker task indicate better inhibitory skills; higher scores in the digit span task indicate larger WMC; and higher scores in the odd-man-out

Table A.2: Tabular overview of modelling accuracy as a function of prosodic condition and length. From top to bottom, (a–b) provide estimate differences between prosodic conditions in medium-length and long sequences, while (c) shows estimate difference between medium-length and long sequences across prosodic conditions.

(a) medium length

comparison	β	SE	Low CrI	High CrI	Evid. Ratio	Post. Prob
neutral vs. rise	-0.11	0.84	-1.46	1.31	0.79	$P(\beta > 0) = 0.44$
neutral vs. fall	0.33	0.80	-0.97	1.66	1.91	$P(\beta > 0) = 0.66$
rise vs. fall	0.43	0.70	-0.70	1.57	2.80	$P(\beta > 0) = 0.74$

(b) long length

comparison	β	SE	Low CrI	High CrI	Evid. Ratio	Post. Prob
neutral vs. rise	-0.57	0.47	-1.34	-0.20	0.13	$P(\beta > 0) = 0.11$
neutral vs. fall	-0.51	0.46	-1.27	0.22	0.14	$P(\beta > 0) = 0.12$
rise vs. fall	0.05	0.57	-0.89	1.00	1.16	$P(\beta > 0) = 0.54$

(c) medium vs. long

comparison	β	SE	Low CrI	High CrI	Evid. Ratio	Post. Prob
neutral vs. rise	0.63	0.48	-0.13	1.43	10.63	$P(\beta > 0) = 0.91$
neutral vs. fall	0.17	0.60	-0.82	1.17	1.53	$P(\beta > 0) = 0.60$
rise vs. fall	-0.21	0.58	-1.18	0.75	0.55	$P(\beta > 0) = 0.35$

Table A.3: Tabular overview of modelling correct response time as a function of prosodic condition and length. From top to bottom, the first two panels provide estimate differences between prosodic conditions in medium-length and long sequences, while the last panel shows estimate difference between medium-length and long sequences across prosodic conditions.

(a) medium length

comparison	β	SE	Low CrI	High CrI	Evid. Ratio	Post. Prob
neutral vs. rise	0.00	0.06	-0.10	0.09	0.89	$P(\beta > 0) = 0.47$
neutral vs. fall	0.00	0.06	-0.10	0.10	0.97	$P(\beta > 0) = 0.49$
rise vs. fall	0.01	0.03	-0.05	0.06	1.31	$P(\beta > 0) = 0.57$

(b) long length

comparison	β	SE	Low CrI	High CrI	Evid. Ratio	Post. Prob
neutral vs. rise	-0.01	0.04	-0.07	0.05	0.55	$P(\beta > 0) = 0.35$
neutral vs. fall	-0.03	0.03	-0.09	0.02	0.18	$P(\beta > 0) = 0.15$
rise vs. fall	-0.02	0.03	-0.08	0.03	0.36	$P(\beta > 0) = 0.26$

(c) medium vs long

comparison	β	SE	Low CrI	High CrI	Evid. Ratio	Post. Prob
neutral vs. rise	0.02	0.03	-0.03	0.08	3.16	$P(\beta > 0) = 0.76$
neutral vs. fall	0.02	0.05	-0.06	0.10	1.67	$P(\beta > 0) = 0.63$
rise vs. fall	-0.01	0.05	-0.09	0.07	0.67	$P(\beta > 0) = 0.40$

A Pilot study – Button press

Table A.4: Mean score performance per participant across the three cognitive tasks.

Participant	Flanker	Odd-man-out	Digit Span
1	1.00	3.25	5
2	0.94	3.33	5
3	0.96	3.72	7
4	0.96	4.65	9
5	0.98	4.52	7
6	1.00	7.57	5
7	1.00	4.17	6
8	0.98	3.32	7
9	0.96	3.42	8
10	0.94	2.90	6
11	0.96	2.80	7
12	0.98	6.32	5
13	0.92	3.78	7
15	0.98	4.97	5
16	0.96	6.60	7
17	0.98	5.15	5
18	0.96	3.50	6
19	1.00	4.47	6
20	1.00	3.37	6
21	0.90	5.29	6
22	0.96	5.22	6
23	1.00	6.52	7
24	0.96	3.31	7
25	1.00	12.41	6
26	0.92	4.42	7
27	0.94	5.01	6
28	0.98	4.06	7
29	0.92	6.09	9
30	0.96	4.47	7

task indicate slower processing speed. A correlation test among the three scores revealed a weak correlation between the flanker task scores and the other two scores. Specifically, a weak positive correlation between flanker and odd-man-out scores was found ($r(1042) = 0.24, p < 0.001$), such that the better the inhibitory skills, the slower the processing speed. In contrast, a weak negative correlation between flanker and digit span scores was observed ($r(1042) = -0.34, p < 0.0001$), such that the better the inhibition, the shorter the span. Given even this weak correlation between two of the cognitive predictors, only one of them was selected for the subsequent analyses, the *flanker scores*, as these indicate inhibitory ability, a cognitive ability highly relevant for the main research question.

Whereas the model on response times provided no reliable evidence for a two-way interaction between prosodic conditions and flanker scores, the model on accuracy provided compelling evidence for such an interaction. Specifically, the model showed that individuals with higher flanker scores achieved reliably lower accuracy in the neutral prosodic condition, but higher accuracy in the rising prosodic condition. Moving to comparisons among prosodic conditions in interaction with flanker scores, the model showed that the higher the flanker score, the better the accuracy in rising and falling conditions, compared to the neutral condition. There were no further differences between the rising and falling conditions. Estimated comparison details for response times are presented in Table A.5, and for accuracy in Table A.6.

A.5.2.3 Speed–accuracy trade-off

Lastly, this section reports on whether there is a speed–accuracy trade-off in the participants' performance. Figure A.3 illustrates the speed–accuracy trade-off using raw values. The zero point (also highlighted by a dashed line) on the y-axis illustrates the onset of the deviant, while the second dashed line represents the end of the sequence. Values between zero and the end of the sequence show correct response times. Negative values show response times before the onset of the deviant. Response times after the offset of the sequence show cases where participants incorrectly did not press the button, and as a result the end of the audio was recorded as the response time. As it can be observed from the overall means, the correct response times do not differ among conditions, an interpretation also supported by the model reported in the confirmatory analyses.

The model on the speed–accuracy trade-off provided compelling evidence for an interaction between speed (response time) and prosodic condition. Specifically, in the neutral prosodic condition, the slower the speed, the better the accuracy, while for the falling condition, the opposite pattern was observed – the

Table A.5: Tabular overview of modelling response times as a function of a two-way interaction between prosodic condition and flanker scores. The table shows mean differences between prosodic conditions with and without interactions (a), as well as differences between prosodic conditions in interaction with attention (b).

(a) Prosody vs prosody:flanker

comparison	β	SE	Low CrI	High CrI	Evid. Ratio	Post. Prob
neutral:flanker	0.03	0.02	-0.01	0.07	6.64	$P(\beta > 0) = 0.87$
rise:flanker	0.01	0.01	-0.01	0.03	2.79	$P(\beta > 0) = 0.74$
fall:flanker	0.00	0.02	-0.03	0.02	0.87	$P(\beta > 0) = 0.47$

(b) prosody:flanker

comparison	β	SE	Low CrI	High CrI	Evid. Ratio	Post. Prob
neutral vs. rise	0.00	0.04	-0.06	0.07	1.11	$P(\beta > 0) = 0.53$
neutral vs. fall	0.00	0.04	-0.06	0.07	1.05	$P(\beta > 0) = 0.51$
rise vs. fall	0.00	0.03	-0.05	0.05	1.15	$P(\beta > 0) = 0.54$

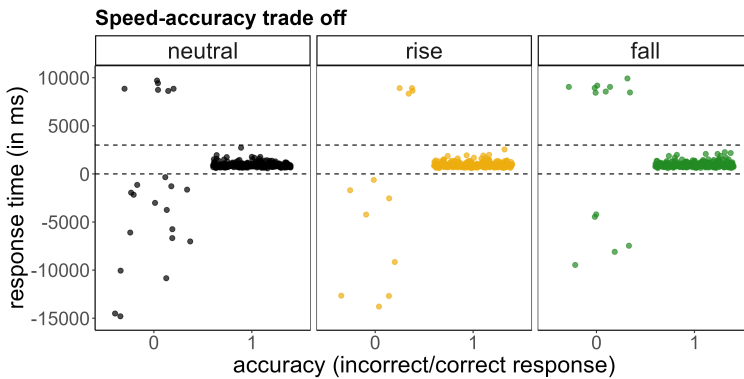


Figure A.3: Speed (y-axis) – accuracy (x-axis) trade-off across prosodic conditions.

Table A.6: Tabular overview of modelling accuracy as a function of a two-way interaction between prosodic condition and flanker scores. The table provides mean differences between prosodic conditions with and without interactions (a), as well as differences between prosodic conditions in interaction with attention (b). Reliable differences are highlighted in bold.

(a) prosody vs prosody:flanker						
comparison	β	SE	Low CrI	High CrI	Evid. Ratio	Post. Prob
neutral:flanker	-0.88	0.38	-1.53	-0.30	130.15	$P(\beta < 0) = \mathbf{0.99}$
rise:flanker	0.71	0.47	-0.06	1.49	14.47	$P(\beta > 0) = \mathbf{0.94}$
fall:flanker	0.42	0.45	-0.32	1.16	4.67	$P(\beta > 0) = 0.82$

(b) prosody:flanker						
comparison	β	SE	Low CrI	High CrI	Evid. Ratio	Post. Prob
neutral vs. rise	2.00	0.77	0.74	3.27	204.13	$P(\beta > 0) = \mathbf{1.00}$
neutral vs. fall	1.57	0.71	0.42	2.76	102.90	$P(\beta > 0) = \mathbf{0.99}$
rise vs. fall	0.43	0.63	-0.62	1.47	3.08	$P(\beta > 0) = 0.76$

slower the speed, the worse the accuracy. This is also shown in Figure A.3. For the neutral condition, most of the incorrect responses occurred on trials where participants pressed the button before the onset of the deviant (i.e., faster speed, less accuracy; slower speed, better accuracy). For the falling condition, the opposite pattern can be observed, that is, most of the incorrect responses resulted from participants not pressing the button at all, as also shown in the model results (slower responses, less accuracy). Finally, the model provided no evidence for an effect of speed on accuracy for the rise condition, or evidence for differences among prosodic conditions in interactions. Estimated comparison details for the speed-accuracy trade-off are presented in Table A.7.

A.6 Discussion

The results of the pilot study point towards an overall ceiling performance across conditions, showing that listeners reliably perceive and detect the deviances in the numeric sequences presented in an explicit behavioural task. At the group

A Pilot study – Button press

Table A.7: Tabular overview of modelling accuracy as a function of speed (response time). The table provides mean differences between prosodic conditions in interaction with speed (a) as well as differences between prosodic conditions with and without interactions (b). Reliable differences are highlighted in bold.

(a) prosody vs prosody:speed

comparison	β	SE	Low CrI	High CrI	Evid. Ratio	Post. Prob
neutral:speed	0.61	0.17	0.34	0.92	Inf	$P(\beta > 0) = \mathbf{1.00}$
rise:speed	0.00	0.27	-0.44	0.45	0.99	$P(\beta > 0) = 0.50$
fall:speed	-0.47	0.24	-0.89	-0.08	41.11	$P(\beta < 0) = \mathbf{0.98}$

(b) prosody:speed

comparison	β	SE	Low CrI	High CrI	Evid. Ratio	Post. Prob
neutral vs. rise	0.02	0.64	-1.05	1.07	1.03	$P(\beta > 0) = 0.50$
neutral vs. fall	0.81	0.59	-0.16	1.79	10.28	$P(\beta > 0) = 0.91$
rise vs. fall	0.71	0.59	-0.17	1.78	10.36	$P(\beta > 0) = 0.91$

level, listener performance did not differ across conditions, indicating that rising intonation did not facilitate the identification of the deviant, or across sequence lengths, showing that there was no benefit of one length over the other. However, the advantage of the rising prosodic condition is subtly in evidence through the speed–accuracy trade-off and individual variability. When examining the relation between speed and accuracy, it is evident that the rising condition is the only condition for which there was no trade-off between speed and accuracy for listeners. On the contrary, there was a trade-off for the other two conditions (faster responses in the neutral condition, and slower responses in the falling condition, lead to worse performance), showing that both neutral and falling intonation challenges listeners in some way in their detection of the deviant. Further, as evinced by the cognitive profiles of the listeners who participated in this pilot study, rising intonation appears to be relevant for successful performance. Specifically, we see that participants with better inhibitory skills showed higher accuracy in the rising condition than average, and that they were better overall in both the rising and falling conditions compared to the neutral one, showing that they need intonation to jolt them out of their inhibition. In particular, the

model indicates that even participants with high inhibitory skills had lower accuracy in the neutral condition, corroborating the finding that neutral intonation impedes the identification of deviants. To conclude, rising intonation appears to be advantageous in detecting the deviant number compared to the neutral and falling conditions, which are more challenging, with the neutral condition being the most likely to involve errors, and the falling condition lying somewhere between the rising and the neutral conditions.

References

- Alain, Claude, David L. Woods & Keith H. Ogawa. 1994. Brain indices of automatic pattern processing. *NeuroReport* 6(1). 140–144. DOI: 10.1097/00001756-199412300-00036.
- Alamia, Andrea, Rufin VanRullen, Emanuele Pasqualotto, André Mouraux & Alexandre Zenon. 2019. Pupil-linked arousal responds to unconscious surprisal. *The Journal of Neuroscience* 39(27). 5369–5376. DOI: 10.1523/JNEUROSCI.3010-18.2019.
- Albert, Aviad. 2023. *A model of sonority based on pitch intelligibility* (Studies in Laboratory Phonology 13). Berlin: Language Science Press. DOI: 10.5281/zenodo.7837176.
- Albert, Aviad, Francesco Cangemi & Martine Grice. 2018. Using periodic energy to enrich acoustic representations of pitch in speech: A demonstration. In *Proceedings of Speech Prosody 2018*, 804–808. International Speech Communication Association. DOI: 10.21437/SpeechProsody.2018-162.
- Albert, Aviad, Francesco Cangemi, Martine Grice & T. Mark Ellison. 2020. *Proper: PROsodic analysis with PERiodic energy*. <https://osf.io/28ea5/>.
- Amiez, Céline & Emmanuel Procyk. 2019. Midcingulate somatomotor and autonomic functions. In Brent A. Vogt (ed.), *Cingulate cortex* (Handbook of Clinical Neurology 166), 53–71. Amsterdam: Elsevier. DOI: 10.1016/B978-0-444-64196-0.00004-2.
- Anwyl-Irvine, Alexander L., Jessica Massonnié, Adam Flitton, Natasha Kirkham & Jo K. Evershed. 2020. Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods* 52(1). 388–407. DOI: 10.3758/s13428-019-01237-x.
- Arvaniti, Amalia. 2022. The autosegmental-metrical model of intonational phonology. In Jonathan Barnes & Stefanie Shattuck-Hufnagel (eds.), *Prosodic theory and practice*, 25–63. The MIT Press. DOI: 10.7551/mitpress/10413.003.0004.
- Assaneo, Florencia M., Johanna M. Rimmele, Joan Orpella, Pablo Ripollés, Ruth de Diego-Balaguer & David Poeppel. 2019. The lateralization of speech-brain coupling is differentially modulated by intrinsic auditory and top-down mechanisms. *Frontiers in Integrative Neuroscience* 13. DOI: 10.3389/fnint.2019.00028.

References

- Bach, Dominik R., Hartmut Schachinger, John G. Neuhoff, Fabrizio Esposito, Francesco D. Salle, Christoph Lehmann, Marcus Herdener, Klaus Scheffler & Erich Seifritz. 2008. Rising sound intensity: An intrinsic warning cue activating the amygdala. *Cerebral Cortex* 18(1). 145–150. DOI: 10.1093/cercor/bhm040.
- Baddeley, Alan. 2003. Working memory: Looking back and looking forward. *Nature Reviews. Neuroscience* 4(10). 829–839. DOI: 10.1038/nrn1201.
- Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278. DOI: 10.1016/j.jml.2012.11.001.
- Baumann, Stefan & Christine T. Röhr. 2015. The perceptual prominence of pitch accent types in German. In *Proceedings of the 18th International Congress of Phonetic Sciences (icphs)*, 1–5. Glasgow, UK.
- Baumann, Stefan & Petra B. Schumacher. 2012. (De-)Accentuation and the processing of information status: Evidence from event-related brain potentials. *Language and Speech* 55(3). 361–381. DOI: 10.1177/0023830911422184.
- Baumann, Stefan & Jürgen Trouvain. 2001. On the prosody of German telephone numbers. In *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech 2001)*, 557–560. International Speech Communication Association. DOI: 10.21437/Eurospeech.2001-149.
- Baumann, Stefan & Bodo Winter. 2018. What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *Journal of Phonetics* 70. 20–38. DOI: 10.1016/j.wocn.2018.05.004.
- Beatty, Jackson. 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin* 91(2). 276–292. DOI: 10.1037/0033-2909.91.2.276.
- Beaumont, J. Graham. 1983. The EEG and task performance: A performance review. In Anthony W. K. Gaillard & Walter Ritter (eds.), *Advances in psychology* (Tutorials in event-related potential research: Endogenous components 10), 385–406. North-Holland. DOI: 10.1016/S0166-4115(08)62049-X.
- Beckman, Mary E. 2012. *Stress and non-Stress accent*. De Gruyter Mouton. DOI: 10.1515/9783110874020.
- Beckman, Mary E. & Jan Edwards. 1994. Articulatory evidence for differentiating stress categories. In Patricia A. Keating (ed.), *Phonological structure and phonetic form: Papers in Laboratory Phonology*, vol. 3, 7–33.
- Beckman, Mary E., Julia Hirschberg & Stefanie Shattuck-Hufnagel. 2005. The original ToBi system and the evolution of the ToBi framework. In Sun-Ah Jun (ed.), *Prosodic typology: The phonology of intonation and phrasing*, 9–54. Oxford University Press. DOI: 10.1093/acprof:oso/9780199249633.003.0002.

- Beckman, Mary E. & Janet B. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology* 3. 255–309. DOI: 10.1017/S095267570000066X.
- Berti, Stefan & Erich Schröger. 2003. Working memory controls involuntary attention switching: Evidence from an auditory distraction paradigm. *The European Journal of Neuroscience* 17(5). 1119–1122. DOI: 10.1046/j.1460-9568.2003.02527.x.
- Bishop, Jason. 2013. Information structural expectations in the perception of prosodic prominence. In Elordieta Gorka & Pilar Prieto (eds.), *Prosody and meaning*, 239–270. De Gruyter Mouton. DOI: 10.1515/9783110261790.239.
- Boersma, Paul & David Weenink. 2024. *Praat: Doing phonetics by computer*. <http://www.praat.org/>.
- Bolinger, Dwight. 1986. *Intonation and its parts: Melody in spoken English*. Stanford University Press.
- Bolinger, Dwight. 1989. *Intonation and its uses: Melody in grammar and discourse*. Stanford University Press.
- Bornkessel-Schlesewsky, Ina & Matthias Schlewsky. 2019. Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Frontiers in Psychology* 10. 298. DOI: 10.3389/fpsyg.2019.00298.
- Bornkessel-Schlesewsky, Ina & Petra B. Schumacher. 2016. Towards a neurobiology of information structure. In Caroline Féry & Shinichiro Ishihara (eds.), *The Oxford handbook of information structure*. Oxford University Press. DOI: 10.1093/oxfordhb/9780199642670.013.22.
- Bornkessel-Schlesewsky, Ina, Isabella Sharrad, Caitlin A. Howlett, Phillip M. Alday, Andrew W. Corcoran, Valeria Bellan, Erica Wilkinson, Reinhold Kliegl, Richard L. Lewis, Steven L. Small & Matthias Schlewsky. 2022. Rapid adaptation of predictive models during language comprehension: Aperiodic EEG slope, individual alpha frequency and idea density modulate individual differences in real-time model updating. *Frontiers in Psychology* 13. DOI: 10.3389/fpsyg.2022.817516.
- Brandreth, Gyles. 1986. *Everyman's word games*. BCA [Book Club Associates - Guild Publishing].
- Braunschweiger, David. 1899. *Die Lehre von der Aufmerksamkeit in der Psychologie des 18. Jahrhunderts*. Leipzig: Haacke.
- Brouwer, Harm & John C. J. Hoeks. 2013. A time and place for language comprehension: Mapping the N400 and the P600 to a minimal cortical network. *Frontiers in Human Neuroscience* 7. DOI: 10.3389/fnhum.2013.00758.
- Bruggeman, Anna. 2020. *Lexical and postlexical prominence in Tashlhiyt Berber and Moroccan Arabic*. Universität zu Köln. (PhD thesis). <https://kups.ub.uni-koeln.de/11189/>.

References

- Bruggeman, Anna, Timo B. Roettger & Martine Grice. 2017. Question word intonation in Tashlhiyt Berber: Is 'high' good enough? *Laboratory Phonology* 8(1). 5. DOI: 10.5334/labphon.79.
- Bürkner, Paul-Christian, Jonah Gabry, Sebastian Weber, Andrew Johnson, Martin Modrak, Hamada S. Badr, Frank Weber, Mattan S. Ben-Shachar, Hayden Rabel, Simon C. Mills & Stephen Wild. 2023. *Brms: Bayesian regression models using 'Stan'*. <https://cran.r-project.org/web/packages/brms/index.html>.
- Cangemi, Francesco. 2015. *Mausmooth*. <https://ifl.phil-fak.uni-koeln.de/sites/linguistik/Phonetik/pdf-publications/2015/cangemi2015mausmooth.pdf>.
- Cangemi, Francesco, Aviad Albert & Martine Grice. 2019. Modelling intonation: Beyond segments and tonal targets. In *Proceedings of the 19th International Congress of Phonetic Sciences*, 572–576. Melbourne, Australia.
- Cangemi, Francesco & Stefan Baumann. 2020. Integrating phonetics and phonology in the study of linguistic prominence. *Journal of Phonetics* 81. 100993. DOI: 10.1016/j.wocn.2020.100993.
- Chandrasekaran, Bharath, Ananthanarayan Krishnan & Jackson T. Gandour. 2009. Sensory processing of linguistic pitch as reflected by the mismatch negativity. *Ear and Hearing* 30(5). 552–558. DOI: 10.1097/AUD.0b013e3181a7e1c2.
- Chen, Aaju. 2003. Language dependence in continuation intonation. In *Proceedings of the 15th International Congress of Phonetic Sciences*, 1069–1072.
- Chobert, Julie, Clément François, Michel Habib & Mireille Besson. 2012. Deficit in the preattentive processing of syllabic duration and VOT in children with dyslexia. *Neuropsychologia* 50(8). 2044–2055. DOI: 10.1016/j.neuropsychologia.2012.05.004.
- Clark, Andy. 2013. Are we predictive engines? Perils, prospects, and the puzzle of the porous perceiver. *Behavioral and Brain Sciences* 36(3). 233–253. DOI: 10.1017/S0140525X12002440.
- Cole, Jennifer, José I. Hualde, Caroline L. Smith, Christopher Eager, Timothy Mahrt & Ricardo Napoleão de Souza. 2019. Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish. *Journal of Phonetics* 75. 113–147. DOI: 10.1016/j.wocn.2019.05.002.
- Cole, Jennifer, Yoonsook Mo & Mark Hasegawa-Johnson. 2010. Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology* 1(2). 425–452. DOI: 10.1515/labphon.2010.022.
- Contier, Friederike, Isabell Wartenburger, Mathias Weymar & Milena Rabovsky. 2024. Are the P600 and P3 ERP components linked to the task-evoked pupillary response as a correlate of norepinephrine activity? *Psychophysiology* 61(7). e14565. DOI: 10.1111/psyp.14565.

- Corbetta, Maurizio & Gordon L. Shulman. 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience* 3(3). 201–215. DOI: 10.1038/nrn755.
- Cowan, Nelson. 1998. *Attention and memory: An integrated framework*. Oxford University Press. DOI: 10.1093/acprof:oso/9780195119107.001.0001.
- Delorme, Arnaud & Scott Makeig. 2004. Eeglab: An open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of Neuroscience Methods* 134(1). 9–21. DOI: 10.1016/j.jneumeth.2003.10.009.
- Demberg, Vera. 2013. Pupillometry: The index of cognitive activity in a dual-task study. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 35. <https://escholarship.org/uc/item/4vf5w6bn>.
- Demberg, Vera & Asad Sayeed. 2016. The frequency of rapid pupil dilations as a measure of linguistic processing difficulty. *PLOS ONE* 11(1). e0146194. DOI: 10.1371/journal.pone.0146194.
- Diascro, Matthew N. & Nathan Brody. 1994. Odd-man-out and intelligence. *Intelligence* 19(1). 79–92. DOI: 10.1016/0160-2896(94)90055-8.
- Dimberg, Ulf. 1990. Facial electromyography and emotional reactions. *Psychophysiology* 27(5). 481–494. DOI: 10.1111/j.1469-8986.1990.tb01962.x.
- Dimitrova, Diana V., Laurie A. Stowe, Gisela Redeker & John C. J. Hoeks. 2012. Less is not more: Neural responses to missing and superfluous accents in context. *Journal of Cognitive Neuroscience* 24(12). 2400–2418. DOI: 10.1162/jocn_a_00302.
- Doeller, Christian F., Bertram Opitz, Axel Mecklinger, Christoph Krick, Wolfgang Reith & Erich Schröger. 2003. Prefrontal cortex involvement in preattentive auditory deviance detection: Neuroimaging and electrophysiological evidence. *NeuroImage* 20(2). 1270–1282. DOI: 10.1016/S1053-8119(03)00389-6.
- Donchin, Emanuel. 1981. Surprise!... Surprise? *Psychophysiology* 18(5). 493–513. DOI: 10.1111/j.1469-8986.1981.tb01815.x.
- Donchin, Emanuel, Walter Ritter & Cheyne McCallum. 1978. Cognitive psychophysiology: The endogenous components of the ERP. *Event-Related Brain Potentials in Man*. 349–441. DOI: 10.1016/B978-0-12-155150-6.X5001-9.
- Duncan, Connie C., Robert J. Barry, John F. Connolly, Catherine Fischer, Patricia T. Michie, Risto Näätänen, John Polich, Ivar Reinvang & Cyma Van Petten. 2009. Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology* 120(11). 1883–1908. DOI: 10.1016/j.clinph.2009.07.045.
- Dynes, Robin. 2017. *The memory handbook: Strategies and activities to aid memory*. Taylor & Francis.

References

- Elliott, Emily M. & Nelson Cowan. 2001. Habituation to auditory distractors in a cross-modal, color–word interference task. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27(3). 654–667. DOI: 10.1037/0278-7393.27.3.654.
- Emmendorfer, Alexandra K., Joao M. Correia, Bernadette M. Jansma, Sonja A. Kotz & Milene Bonte. 2020. ERP mismatch response to phonological and temporal regularities in speech. *Scientific Reports* 10(1). 9917. DOI: 10.1038/s41598-020-66824-x.
- Eriksen, Barbara A. & Charles W. Eriksen. 1974. Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics* 16(1). 143–149. DOI: 10.3758/BF03203267.
- Escera, Carles, Kimmo Alho, István Winkler & Risto Näätänen. 1998. Neural mechanisms of involuntary attention to acoustic novelty and change. *Journal of Cognitive Neuroscience* 10(5). 590–604. DOI: 10.1162/089892998562997.
- Fothergill, Alastair & Andy Byatt. 2003. *Deep blue*. English. Germany/United Kingdom.
- Fraundorf, Scott H., Duane G. Watson & Aaron S. Benjamin. 2010. Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of memory and language* 63(3). 367–386. DOI: 10.1016/j.jml.2010.06.004.
- Frazier, Lyn, Katy Carlson & Charles Clifton. 2006. Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences* 10(6). 244–249. DOI: 10.1016/j.tics.2006.04.002.
- Frearson, Warwick & Hans J. Eysenck. 1986. Intelligence, reaction time (RT) and a new “odd-man-out” RT paradigm. *Personality and Individual Differences* 7(6). 807–817. DOI: 10.1016/0191-8869(86)90079-6.
- Friederici, Angela D., Yunhua Wang, Christoph S. Herrmann, Burkhard Maess & Ulrich Oertel. 2000. Localization of early syntactic processes in frontal and temporal cortical areas: A magnetoencephalographic study. *Human Brain Mapping* 11(1). 1–11. DOI: 10.1002/1097-0193(200009)11:1<1::AID-HBM10>3.0.CO;2-B.
- Frischkorn, Gidon T., Oliver Wilhelm & Klaus Oberauer. 2022. Process-oriented intelligence research: A review from the cognitive perspective. *Intelligence* 94. 101681. DOI: 10.1016/j.intell.2022.101681.
- Friston, Karl. 2010. The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience* 11(2). 127–138. DOI: 10.1038/nrn2787.
- Friston, Karl. 2018. Does predictive coding have a future? *Nature Neuroscience* 21(8). 1019–1021. DOI: 10.1038/s41593-018-0200-7.

- Gandour, Jackson, Yunxia Tong, Donald Wong, Thomas Talavage, Mario Dzemidzic, Yisheng Xu, Xiaojian Li & Mark Lowe. 2004. Hemispheric roles in the perception of speech prosody. *NeuroImage* 23(1). 344–357. DOI: 10.1016/j.neuroimage.2004.06.004.
- Gati, Itamar & Gershon Ben-Shakhar. 1990. Novelty and significance in orientation and habituation: A feature-matching approach. *Journal of Experimental Psychology. General* 119(3). 251–263. DOI: 10.1037//0096-3445.119.3.251.
- Gibson, James J. 1966. *The senses considered as perceptual systems*. Houghton Mifflin.
- Grabe, Esther. 1998. *Comparative intonational phonology: English and German*. MPI Series in Psycholinguistics 7. Wageningen, Ponsen en Looien. (PhD thesis).
- Graham, Frances K. & Rachel K. Clifton. 1966. Heart-rate change as a component of the orienting response. *Psychological Bulletin* 65(5). 305–320. DOI: 10.1037/h0023258.
- Gregory, Richard L. 1970. *The intelligent eye*. London: Weidenfeld & Nicolson.
- Grice, Martine. 2022a. Autosegmental-metrical phonology: Unpacking the boxes. *Zeitschrift für Sprachwissenschaft* 41(2). 393–411. DOI: 10.1515/zfs-2022-2002.
- Grice, Martine. 2022b. Commentary on Chapter 1: Introducing flexibility into autosegmental-metrical phonology. In Stefanie Shattuck-Hufnagel & Jonathan Barnes (eds.), *Prosodic theory and practice*, 64–75. 10.7551/mitpress/10413.003.0016. The MIT Press.
- Grice, Martine & Stefan Baumann. 2007. An introduction to intonation: Functions and models. In Jürgen Trouvain & Ulrike Gut (eds.), *Proceedings of non-native prosody. Phonetic description and teaching practice*, 25–52. De Gruyter (= Trends in Linguistics. Studies & Monographs [TiLSM] 186). <https://api.semanticscholar.org/CorpusID:10691377>.
- Grice, Martine, Stefan Baumann & Ralf Benz Müller. 2005. German intonation in autosegmental-metrical phonology. In Sun-Ah Jun (ed.), *Prosodic typology: The phonology of intonation and phrasing*, 55–83. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199249633.003.0003.
- Grice, Martine & Frank Kügler. 2021. Prosodic prominence: A cross-linguistic perspective. *Language and Speech* 64(2). 253–260. DOI: 10.1177/00238309211015768.
- Grice, Martine, Simon Ritter, Henrik Niemann & Timo B. Roettger. 2017. Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics. Mechanisms of regulation in speech* 64. 90–107. DOI: 10.1016/j.wocn.2017.03.003.

References

- Grice, Martine, Michelina Savino, Maria Lialiou, Aviad Albert, Caterina Ventura & Petra B. Schumacher. 2025. The perception of post-focal prominence is (also) a matter of linguistic structure. *Journal of the International Phonetic Association*. 1–22. DOI: 10.1017/S0025100325100479.
- Grice, Martine, Michelina Savino, Petra B. Schumacher, Christine Tanja Röhr & T. Mark Ellison. 2024. Rises on pitch accents and edge tones affect serial recall performance at item and domain levels. *Laboratory Phonology* 15(1). DOI: <https://doi.org/10.16995/labphon.10473>.
- Grice, Martine, Alexandra Vella & Anna Bruggeman. 2019. Stress, pitch accent, and beyond: Intonation in Maltese questions. *Journal of Phonetics* 76. 100913. DOI: 10.1016/j.wocn.2019.100913.
- Grice, Martine, Simon Wehrle, Martina Krüger, Malin Spaniol, Francesco Cangemi & Kai Vogeley. 2023. Linguistic prosody in autism spectrum disorder—An overview. *Language and Linguistics Compass* 17(5). e12498. DOI: 10.1111/lnc3.12498.
- Gussenhoven, Carlos. 2004. *The phonology of tone and intonation* (Research Surveys in Linguistics). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511616983.
- Gussenhoven, Carlos, Bruno H. Repp, Toni Rietveld, Hans Hardy Rump & Jacques Terken. 1997. The perceptual prominence of fundamental frequency peaks. *The Journal of the Acoustical Society of America* 102(5). 3009–3022. <https://pubs.aip.org/asa/jasa/article-abstract/102/5/3009/558117>.
- Gussenhoven, Carlos & Toni Rietveld. 1988. Fundamental frequency declination in Dutch: Testing three hypotheses. *Journal of Phonetics* 16(3). 335–369. DOI: 10.1016/S0095-4470(19)30509-1.
- Hart, John T., René Collier & Antonie Cohen. 1990. *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press.
- Hasting, Anna S., Sonja A. Kotz & Angela D. Friederici. 2007. Setting the stage for automatic syntax processing: The mismatch negativity as an indicator of syntactic priming. *Journal of Cognitive Neuroscience* 19(3). 386–400. DOI: 10.1162/jocn.2007.19.3.386.
- Hatfield, Gary. 1998. Attention in early scientific psychology. In Richard D. Wright (ed.), *Visual attention*, 3–25. Oxford University Press.
- Häuser, Katja I., Vera Demberg & Jutta Kray. 2019. Effects of aging and dual-task demands on the comprehension of less expected sentence continuations: Evidence from pupillometry. *Frontiers in Psychology* 10. DOI: 10.3389/fpsyg.2019.00709.

- Heitz, Richard P. & Randall W. Engle. 2007. Focusing the spotlight: Individual differences in visual attention control. *Journal of Experimental Psychology: General* 136(2). 217–240. DOI: 10.1037/0096-3445.136.2.217.
- Hickok, Gregory & David Poeppel. 2007. The cortical organization of speech processing. *Nature Reviews Neuroscience* 8(5). 393–402. DOI: 10.1038/nrn2113.
- Himmelman, Nikolaus P. & Beatrice Primus. 2015. Prominence beyond prosody – A first approximation. In *Proceedings of Viterbo: DISUCOM press*, 38–58.
- Hommel, Bernhard, Craig S. Chapman, Paul Cisek, Heather F. Neyedli, Joo-Hyun Song & Timothy N. Welsh. 2019. No one knows what attention is. *Attention, Perception & Psychophysics* 81(7). 2288–2303. DOI: 10.3758/s13414-019-01846-w.
- Hsu, Chun-Hsien, Jonathan P. Evans & Chia-Ying Lee. 2015. Brain responses to spoken F0 changes: Is H special? *Journal of Phonetics* 51. 82–92. DOI: 10.1016/j.wocn.2015.02.003.
- Hsu, Yi-Fang, Chia-An Tu, Yuchun Chen & Huei-Mei Liu. 2023. The mismatch negativity to abstract relationship of tone pairs is independent of attention. *Scientific Reports* 13(1). 9839. DOI: 10.1038/s41598-023-37131-y.
- Huettig, Falk. 2015. Four central questions about prediction in language processing. *Brain Research. Predictive and Attentive Processing in Perception and Action* 1626. 118–135. DOI: 10.1016/j.brainres.2015.02.014.
- Hughes, Robert W. 2014. Auditory distraction: A duplex-mechanism account. *PsyCh Journal* 3(1). 30–41. DOI: 10.1002/pchj.44.
- Hughes, Robert W. & Dylan M. Jones. 2003. A negative order-repetition priming effect: Inhibition of order in unattended auditory sequences? *Journal of Experimental Psychology: Human Perception and Performance* 29(1). 199–218. DOI: 10.1037/0096-1523.29.1.199.
- Hughes, Robert W. & Dylan M. Jones. 2005. The impact of order incongruence between a task-irrelevant auditory sequence and a task-relevant visual sequence. *Journal of Experimental Psychology: Human Perception and Performance* 31(2). 316–327. DOI: 10.1037/0096-1523.31.2.316.
- Hughes, Robert W., François Vachon & Dylan M. Jones. 2007. Disruption of short-term memory by changing and deviant sounds: Support for a duplex-mechanism account of auditory distraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33(6). 1050–1061. DOI: 10.1037/0278-7393.33.6.1050.
- James, William. 1890. *The principles of psychology*. London, England: Dover Publications.
- Jesse, Harris A. & Sun-Ah Jun. 2019. Using pupillometry to assess prosodic alignment in language. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sci-*

References

- ences, 2926–2930. Melbourne: Australasian Speech Science & Technology Association Inc.
- Johansson, Birger & Christian Balkenius. 2018. A computational model of pupil dilation. *Connection Science* 30(1). 5–19. DOI: 10.1080/09540091.2016.1271401.
- Joshi, Siddhartha & Joshua I. Gold. 2020. Pupil size as a window on neural substrates of cognition. *Trends in Cognitive Sciences* 24(6). 466–480. DOI: 10.1016/j.tics.2020.03.005.
- Jun, Sun-Ah. 2005a. Introduction. In Sun-Ah Jun (ed.), *Prosodic typology: The phonology of intonation and phrasing*, 1–8. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199249633.003.0001.
- Jun, Sun-Ah (ed.). 2005b. *Prosodic typology: The phonology of intonation and phrasing*. Oxford University Press. DOI: 10.1093/acprof:oso/9780199249633.001.0001.
- Jun, Sun-Ah (ed.). 2014a. *Prosodic typology II: The phonology of intonation and phrasing*. Oxford: Oxford University Press.
- Jun, Sun-Ah. 2014b. Prosodic typology: By prominence type, word prosody, and macro-rhythm. In Sun-Ah Jun (ed.), *Prosodic typology II: The phonology of intonation and phrasing*, 520–539. Oxford, New York: Oxford University Press. DOI: 10.1093/acprof:oso/9780199567300.003.0017.
- Kakouros, Sofoklis, Nelli Salminen & Okko Räsänen. 2018. Making predictable unpredictable with style: Behavioral and electrophysiological evidence for the critical role of prosodic expectations in the perception of prominence in speech. *Neuropsychologia* 109. 181–199. DOI: 10.1016/j.neuropsychologia.2017.12.011.
- Kember, Heather, Jiyoun Choi, Jenny Yu & Anne Cutler. 2021. The processing of linguistic prominence. *Language and Speech* 64(2). 413–436. DOI: 10.1177/0023830919880217.
- Keye, Doris, Oliver Wilhelm, Klaus Oberauer & Don van Ravenzwaaij. 2009. Individual differences in conflict-monitoring: Testing means and covariance hypothesis about the Simon and the Eriksen Flanker task. *Psychological Research* 73(6). 762–776. DOI: 10.1007/s00426-008-0188-9.
- Knight, Rachael-Anne. 2008. The shape of nuclear falls and their effect on the perception of pitch and prominence: Peaks vs. plateaux. *Language and Speech* 51(3). 223–244. DOI: 10.1177/0023830908098541.
- Koch, Xaver & Katharina Spalek. 2021. Contrastive intonation effects on word recall for information-structural alternatives across the sexes. *Memory & Cognition* 49(7). 1312–1333. DOI: 10.3758/s13421-021-01174-1.
- Koffka, Kurt. 1935. *Principles of Gestalt psychology*. Oxford, England: Harcourt, Brace.
- Kohler, Klaus & Robert Gartenberg. 1991. The perception of accents: F0 peak height versus F0 peak position. In *Proceedings of AIPUK*, vol. 25, 219–242.

- Kreegipuu, Kairi & Risto Näätänen. 2009. The mismatch negativity as an index of different forms of memory in audition. In Lars Nyberg Lars Bäckman (ed.), *Memory, aging and the brain*. Psychology Press.
- Kyröläinen, Aki-Juhani, Vincent Porretta, Jacolien van Rij & Juhani Järvikivi. 2020. *PupilPre: Preprocessing pupil size data*. <https://cran.r-project.org/web/packages/PupilPre/index.html>.
- Ladd, D. Robert. 2008. *Intonational phonology*. 2nd edn. (Cambridge Studies in Linguistics). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511808814.
- Ladd, D. Robert & Rachel Morton. 1997. The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics* 25(3). 313–342. DOI: 10.1006/jpho.1997.0046.
- Lauchlan, Fraser & Donna Carrigan. 2013. *Improving learning through dynamic assessment: A practical classroom resource*. Jessica Kingsley Publishers.
- Leiner, Dominik J. 2024. *SoSci Survey*. www.sosicisurvey.de.
- Li, Weijun, Nali Deng, Yufang Yang & Lin Wang. 2018. Process focus and accentuation at different positions in dialogues: An ERP study. *Language, Cognition and Neuroscience* 33(2). 255–274. DOI: 10.1080/23273798.2017.1387278.
- Li, Xiaoqing & Yiya Chen. 2018. Unattended processing of hierarchical pitch variations in spoken sentences. *Brain and Language* 183. 21–31. DOI: 10.1016/j.bandl.2018.05.004.
- Lialiou, Maria, Aviad Albert, Alexandra Vella & Martine Grice. 2021. Periodic energy mass on head and edge tones in Maltese wh-constructions. In *Proceedings of the 1st International Conference on Tone and Intonation (TAI)*, 161–165. International Speech Communication Association. DOI: 10.21437/TAI.2021-33.
- Lialiou, Maria, Anna Bruggeman, Alexandra Vella, Sarah Grech, Petra Schumacher & Martine Grice. 2023. Word-level prominence and “stress deafness” in Maltese-English Bilinguals. In Radek Skarnitzl & Jan Volín (eds.), *Proceedings of the 20th International Congress of Phonetic Sciences*, 132–136. Guarant International.
- Lialiou, Maria, Martine Grice, Christine T. Röhr & Petra B. Schumacher. 2024. Auditory processing of intonational rises and falls in German: Rises are special in attention orienting. *Journal of Cognitive Neuroscience*. 1–24. DOI: 10.1162/jocn_a_02129.
- Lialiou, Maria, Jesse A. Harris, Martine Grice & Petra B. Schumacher. 2024. Attention allocation to deviants with intonational rises and falls: Evidence from pupillometry. In *Proceedings of the Annual Conference of Cognitive Science Society*, vol. 46, 667–674. <https://escholarship.org/uc/item/68q9c2r8>.

References

- Liao, Hsin-I, Makoto Yoneya, Shunsuke Kidani, Makio Kashino & Shigeto Furukawa. 2016. Human pupillary dilation response to deviant auditory stimuli: Effects of stimulus properties and voluntary attention. *Frontiers in Neuroscience* 10. DOI: 10.3389/fnins.2016.00043.
- Liberman, Mark Yoffe. 1975. *The intonational system of English*. Massachusetts Institute of Technology. (PhD thesis). <https://dspace.mit.edu/handle/1721.1/27376>.
- Liu, Min, Yiya Chen & Niels O. Schiller. 2016. Online processing of tone and intonation in Mandarin: Evidence from ERPs. *Neuropsychologia* 91. 307–317. DOI: 10.1016/j.neuropsychologia.2016.08.025.
- Lorenzen, Janne, Simon Roessig & Stefan Baumann. 2022. Information status and tonal context jointly modulate prosodic prominence relations in German. In *Proceedings of Speech Prosody 2022*, 7–11. DOI: 10.21437/SpeechProsody.2022-2.
- Lorenzen, Janne, Simon Roessig & Stefan Baumann. 2023. Redundancy and individual variability in the prosodic marking of information status in German. In Radek Skarnitzl & Jan Volín (eds.), *Proceedings of the 20th International Congress of Phonetic Sciences*, 1320–1324. Guarant International.
- Lorenzen, Janne, Simon Roessig & Stefan Baumann. 2024. Paradigmatic and syntagmatic effects of information status on prosodic prominence – Evidence from an interactive web-based production experiment in German. *Frontiers in Psychology* 15. DOI: 10.3389/fpsyg.2024.1296933.
- Luck, Steven J. 2014. *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.
- Lynn, Richard. 1966. *Attention, arousal, and the orientation reaction*. 1st ed. (International series of monographs in experimental psychology v. 3). Oxford, New York: Pergamon Press.
- Macdonald, Margaret & Kenneth Campbell. 2011. Effects of a violation of an expected increase or decrease in intensity on detection of change within an auditory pattern. *Brain and Cognition* 77(3). 438–445. DOI: 10.1016/j.bandc.2011.08.014.
- Maess, Burkhard, Erich Schröger & Andreas Widmann. 2016a. High-pass filters and baseline correction in M/EEG analysis-continued discussion. *Journal of Neuroscience Methods* 266. 171–172. DOI: 10.1016/j.jneumeth.2016.01.016.
- Maess, Burkhard, Erich Schröger & Andreas Widmann. 2016b. High-pass filters and baseline correction in M/EEG analysis. Commentary on: “How inappropriate high-pass filters can produce artefacts and incorrect conclusions in ERP studies of language and cognition”. *Journal of Neuroscience Methods* 266. 164–165. DOI: 10.1016/j.jneumeth.2015.12.003.

- Maltzman, Irving. 1971. The orienting reflex and thinking as determiners of conditioning and generalization to words. In Howard H. Kendler & Janet T. Spence (eds.), *Essays in neobehaviorism: A memorial volume to Kenneth W. Spence*. New York, NY: Appleton-Century-Crofts.
- Maltzman, Irving. 1979. Orienting reflexes and classical conditioning in humans. In H. D. Kimmel, E. H. van Olst & J. F. Orlebeke (eds.), *The orienting reflex in humans*, 323–352. London: Routledge.
- Maltzman, Irving & Mary P. Mandell. 1968. The orienting reflex as a predictor of learning and performance. *Journal of Experimental Research in Personality* 3(2). 99–106.
- Maltzman, Irving & David C. Raskin. 1965. Effects of individual differences in the orienting reflex on conditioning and complex processes. *Journal of Experimental Research in Personality* 1(1). 1–16.
- Marois, Alexandre, Katherine Labonté, Mark Parent & François Vachon. 2018. Eyes have ears: Indexing the orienting response to sound using pupillometry. *International Journal of Psychophysiology* 123. 152–162. DOI: 10.1016/j.ijpsycho.2017.09.016.
- Marois, Alexandre, John E. Marsh & François Vachon. 2019. Is auditory distraction by changing-state and deviant sounds underpinned by the same mechanism? Evidence from pupillometry. *Biological Psychology* 141. 64–74. DOI: 10.1016/j.biopsycho.2019.01.002.
- Mathôt, Sebastiaan, Daniel Schreij & Jan Theeuwes. 2012. OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods* 44(2). 314–324. DOI: 10.3758/s13428-011-0168-7.
- Milner, Peter M. 1974. A model for visual shape recognition. *Psychological Review* 81(6). 521–535. DOI: 10.1037/h0037149.
- Müller, Hermann J. & Patrick M. Rabbitt. 1989. Reflexive and voluntary orienting of visual attention: Time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance* 15(2). 315–330. DOI: 10.1037/0096-1523.15.2.315.
- Näätänen, Risto. 1979. Orienting and evoked potentials. In H. D. Kimmel, E. H. van Olst & J. F. Orlebeke (eds.), *The orienting reflex in humans*, 61–75. London: Routledge.
- Näätänen, Risto. 1990. The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences* 13(2). 201–288. DOI: 10.1017/S0140525X00078407.
- Näätänen, Risto. 1992. *Attention and brain function*. Hillsdale: Lawrence Erlbaum Associates.

References

- Näätänen, Risto. 2001. The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology* 38(1). 1–21. DOI: 10.1111/1469-8986.3810001.
- Näätänen, Risto, Anthony W. K. Gaillard & Sirkka Mäntysalo. 1978. Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica* 42(4). 313–329. DOI: 10.1016/0001-6918(78)90006-9.
- Näätänen, Risto, Anthony W. K. Gaillard & Sirkka Mäntysalo. 1980. Brain potential correlates of voluntary and involuntary attention. *Progress in Brain Research* 54. 343–348. DOI: 10.1016/S0079-6123(08)61645-3.
- Näätänen, Risto, Teija Kujala & Gregory Light. 2019. *The mismatch negativity: A window to the brain*. Oxford: Oxford University Press. DOI: 10.1093/oso/9780198705079.001.0001.
- Näätänen, Risto, Teija Kujala & István Winkler. 2011. Auditory processing that leads to conscious perception: A unique window to central auditory processing opened by the mismatch negativity and related responses. *Psychophysiology* 48(1). 4–22. DOI: 10.1111/j.1469-8986.2010.01114.x.
- Näätänen, Risto, Petri Paavilainen, Kimo Alho, Kalevi Reinikainen & Mikko Sams. 1989. Do event-related potentials reveal the mechanism of the auditory sensory memory in the human brain? *Neuroscience Letters* 98(2). 217–221. DOI: 10.1016/0304-3940(89)90513-2.
- Näätänen, Risto, Petri Paavilainen, Teemu Rinne & Kimmo Alho. 2007. The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology* 118(12). 2544–2590. DOI: 10.1016/j.clinph.2007.04.026.
- Näätänen, Risto, Petri Paavilainen, Hannu Tiitinen, Deming Jiang & Kimmo Alho. 1993. Attention and mismatch negativity. *Psychophysiology* 30(5). 436–450. DOI: 10.1111/j.1469-8986.1993.tb02067.x.
- Näätänen, Risto, Mari Tervaniemi, Elyse Sussman, Petri Paavilainen & István Winkler. 2001. “Primitive intelligence” in the auditory cortex. *Trends in Neurosciences* 24(5). 283–288. DOI: 10.1016/s0166-2236(00)01790-2.
- Näätänen, Risto & István Winkler. 1999. The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin* 125(6). 826–859. DOI: 10.1037/0033-2909.125.6.826.
- Niebuhr, Oliver. 2009. F0-based rhythm effects on the perception of local syllable prominence. *Phonetica* 66(1-2). 95–112. DOI: 10.1159/000208933.
- Nieuwenhuis, Sander, Gary Aston-Jones & Jonathan D. Cohen. 2005. Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin* 131(4). 510–532. DOI: 10.1037/0033-2909.131.4.510.

- Nieuwenhuis, Sander, Eco J. De Geus & Gary Aston-Jones. 2011. The anatomical and functional relationship between the P3 and autonomic components of the orienting response. *Psychophysiology* 48(2). 162–175. DOI: 10.1111/j.1469-8986.2010.01057.x.
- Oldfield, Richard C. 1971. The assessment and analysis of handedness: The Edinburgh inventory. eng. *Neuropsychologia* 9(1). 97–113. DOI: 10.1016/0028-3932(71)90067-4.
- Orrico, Riccardo, Stella Gryllia, Na Hu, Jiseung Kim & Amalia Arvaniti. 2024. Prosodic prominence in Greek: Methodological and theoretical considerations. In *Proceedings of Speech Prosody 2024*.
- Ou, Shu-chen & Zhe-chen Guo. 2021. The language-specific use of fundamental frequency rise in segmentation of an artificial language: Evidence from listeners of Taiwanese Southern Min. *Language and Speech* 64(2). 437–466. DOI: 10.1177/0023830919886604.
- Paavilainen, Petri. 2013. The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: A review. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology* 88(2). 109–123. DOI: 10.1016/j.ijpsycho.2013.03.015.
- Pavlov, Ivan Petrovich. 1927. *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. Oxford: Oxford Univ. Press.
- Pavlov, Ivan Petrovich. 1955. *Selected works*. Moscow: Foreign Languages Publishing House.
- Peters, Jörg. 2018. Phonological and semantic aspects of German intonation. *Linguistik Online* 88(1). DOI: 10.13092/lo.88.4191.
- Petersen, Steven E. & Michael I. Posner. 2012. The attention system of the human brain: 20 years after. *Annual Review of Neuroscience* 35. 73–89. DOI: 10.1146/annurev-neuro-062111-150525.
- Pierrehumbert, Janet B. 1980. *The phonology and phonetics of English intonation*. MIT. (PhD thesis).
- Polich, John. 1986. Normal variation of P300 from auditory stimuli. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section* 65(3). 236–240. DOI: 10.1016/0168-5597(86)90059-6.
- Polich, John. 2007. Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology* 118(10). 2128–2148. DOI: 10.1016/j.clinph.2007.04.019.
- Posner, Michael I. 1978. *Chronometric explorations of mind*. Oxford: Lawrence Erlbaum.
- Posner, Michael I. 1980. Orienting of attention. *Quarterly Journal of Experimental Psychology* 32(1). 3–25. DOI: 10.1080/00335558008248231.

References

- Posner, Michael I. & Steven E. Petersen. 1990. The attention system of the human brain. *Annual Review of Neuroscience* 13. 25–42. DOI: 10.1146/annurev.ne.13.030190.000325.
- Quasthoff, Uwe & Matthias Richter. 2005. Projekt Deutscher Wortschatz. *Babylonia*. 33–35.
- R Core Team. 2023. *R: A language and environment for statistical computing*. <https://www.R-project.org/>.
- Reisberg, Daniel, Jonathan Baron & Deborah G. Kemler. 1980. Overcoming Stroop interference: The effects of practice on distractor potency. *Journal of Experimental Psychology: Human Perception and Performance* 6(1). 140–150. DOI: 10.1037/0096-1523.6.1.140.
- Ren, Guiqin, Yufang Yang & Xiaoqing Li. 2009. Early cortical processing of linguistic pitch patterns as revealed by the mismatch negativity. *Neuroscience* 162(1). 87–95. DOI: 10.1016/j.neuroscience.2009.04.021.
- Repp, Magdalena & Petra B. Schumacher. 2023. What naturalistic stimuli tell us about pronoun resolution in real-time processing. *Frontiers in Artificial Intelligence* 6. DOI: 10.3389/frai.2023.1058554.
- Rietveld, Toni & Carlos Gussenhoven. 1985. On the relation between pitch excursion size and prominence. *Journal of Phonetics* 13(3). 299–308. DOI: 10.1016/S0095-4470(19)30761-2.
- Rinne, Teemu, Alexander Degerman & Kimmo Alho. 2005. Superior temporal and inferior frontal cortices are activated by infrequent sound duration decrements: An fMRI study. *NeuroImage* 26(1). 66–72. DOI: 10.1016/j.neuroimage.2005.01.017.
- Rinne, Teemu, Anna Säkkä, Alexander Degerman, Erich Schröger & Kimmo Alho. 2006. Two separate mechanisms underlie auditory change detection and involuntary control of attention. *Brain Research* 1077(1). 135–143. DOI: 10.1016/j.brainres.2006.01.043.
- Roessig, Simon. 2021. *Categoriality and continuity in prosodic prominence* (Studies in Laboratory Phonology 10). Berlin: Language Science Press. DOI: 10.5281/zenodo.4121875.
- Roessig, Simon & Doris Mücke. 2019. Modeling dimensions of prosodic prominence. *Frontiers in Communication* 4. DOI: 10.3389/fcomm.2019.00044.
- Roessig, Simon, Doris Mücke & Martine Grice. 2019. The dynamics of intonation: Categorical and continuous variation in an attractor-based model. *PLOS ONE* 14(5). e0216859. DOI: 10.1371/journal.pone.0216859.
- Roettger, Timo B. & Kim Rimland. 2020. Listeners' adaptation to unreliable intonation is speaker-sensitive. *Cognition* 204. 104372. DOI: 10.1016/j.cognition.2020.104372.

- Röhr, Christine T., Stefan Baumann & Martine Grice. 2022. The influence of expectations on tonal cues to prominence. *Journal of Phonetics* 94. 101–174. DOI: 10.1016/j.wocn.2022.101174.
- Röhr, Christine T., Stefan Baumann, Petra B. Schumacher & Martine Grice. 2020. Perceptual prominence of accent types and the role of expectations. In *Proceedings of Speech Prosody 2020*, 366–370. International Speech Communication Association. DOI: 10.21437/SpeechProsody.2020-75.
- Röhr, Christine T., Ingmar Brilmayer, Stefan Baumann, Martine Grice & Petra B. Schumacher. 2021. Signal-driven and expectation-driven processing of accent types. *Language, Cognition and Neuroscience* 36(1). 33–59. DOI: 10.1080/23273798.2020.1779324.
- Röhr, Christine T., Michelina Savino & Martine Grice. 2022. The effect of intonational rises on serial recall in German. In *Proceedings of Speech Prosody 2022*, 759–763. DOI: 10.21437/SpeechProsody.2022-154.
- Rosenberg, Andrew, Erica Cooper, Rivka Levitan & Julia Hirschberg. 2012. Cross-language prominence detection. In *Proceedings of Speech Prosody 2012*, 278–281. DOI: 10.21437/SpeechProsody.2012-71.
- Ruchkin, Daniel S., Ray Johnson Jr., Howard L. Canoune, Walter Ritter & Muriel Hammer. 1990. Multiple sources of P3b associated with different types of information. *Psychophysiology* 27(2). 157–176. DOI: 10.1111/j.1469-8986.1990.tb00367.x.
- SanMiguel, Iria, María-José Corral & Carles Escera. 2008. When loading working memory reduces distraction: Behavioral and electrophysiological evidence from an auditory-visual distraction paradigm. *Journal of Cognitive Neuroscience* 20(7). 1131–1145. DOI: 10.1162/jocn.2008.20078.
- Savino, Michelina, Bodo Winter, Andrea Bosco & Martine Grice. 2020. Intonation does aid serial recall after all. *Psychonomic Bulletin & Review* 27(2). 366–372. DOI: 10.3758/s13423-019-01708-4.
- Scheepers, Christoph, Sibylle Mohr, Martin H. Fischer & Andrew M. Roberts. 2013. Listening to limericks: A pupillometry investigation of perceivers' expectancy. *PLOS ONE* 8(9). e74986. DOI: 10.1371/journal.pone.0074986.
- Schumacher, Petra B., Jana Backhaus & Manuel Dangel. 2015. Backward- and forward-looking potential of anaphors. *Frontiers in Psychology* 6. 17–46. DOI: 10.3389/fpsyg.2015.01746.
- Schumacher, Petra B. & Stefan Baumann. 2010. Pitch accent type affects the N400 during referential processing. *NeuroReport* 21(9). 618–622. DOI: 10.1097/WNR.0b013e328339874a.

References

- Seidl, Amanda & Elizabeth K. Johnson. 2006. Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental Science* 9(6). 565–573. DOI: 10.1111/j.1467-7687.2006.00534.x.
- Shattuck-Hufnagel, Stefanie. 2020. The role of phrase-level prosody in speech production. In Carlos Gussenhoven & Aoju Chen (eds.), *The Oxford handbook of language prosody*, 522–538. Oxford: Oxford University Press.
- Shattuck-Hufnagel, Stefanie & Alice E. Turk. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25(2). 193–247. DOI: 10.1007/BF01708572.
- Smith, Rachel & Tamara Rathcke. 2020. Dialectal phonology constrains the phonetics of prominence. *Journal of Phonetics* 78. 100934. DOI: 10.1016/j.wocn.2019.100934.
- Sokolov, Evgenii N. 1963a. *Perception and the conditioned reflex*. Oxford: Pergamon Press.
- Sokolov, Evgenii Nikolaevich. 1960. Neuronal models and the orienting reflex. In M. A. Brazier (ed.), *The central nervous system and behavior*, 187–276. Josiah Macy Jr. Foundation.
- Sokolov, Evgenii Nikolaevich. 1963b. Higher nervous functions: The orienting reflex. *Annual Review of Physiology* 25(1). 545–580. DOI: 10.1146/annurev.ph.25.030163.002553.
- Sörqvist, Patrik, John E. Marsh & Anatole Nöstl. 2013. High working memory capacity does not always attenuate distraction: Bayesian evidence in support of the null hypothesis. *Psychonomic Bulletin & Review* 20(5). 897–904. DOI: 10.3758/s13423-013-0419-y.
- Sóskuthy, Márton. 2017. *Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction*. DOI: 10.48550/arXiv.1703.05339.
- Spilisbury, Georgina, Lazar Stankov & Richard Roberts. 1990. The effect of a test's difficulty on its correlation with intelligence. *Personality and Individual Differences* 11(10). 1069–1077. DOI: 10.1016/0191-8869(90)90135-E.
- Strauch, Christoph, Chin-An Wang, Wolfgang Einhäuser, Stefan van der Stigchel & Marnix Naber. 2022. Pupillometry as an integrated readout of distinct attentional networks. *Trends in Neurosciences* 45(8). 635–647. DOI: 10.1016/j.tins.2022.05.003.
- Streefkerk, Barbertje M. 2002. *Prominence. Acoustic and lexical/syntactic correlates*. UtrechtLOT. (PhD thesis). <https://dare.uva.nl/search?identifier=44dd581d-7c95-442b-96fe-9a6a1e19c1f3>.
- Sullivan, Joseph W. & Frances Degen Horowitz. 1983. The effects of intonation on infant attention: The role of the rising intonation contour. *Journal of Child Language* 10(3). 521–534. DOI: 10.1017/S0305000900005341.

- Sussman, Elyse S., Sufen Chen, Jon Sussman-Fort & Elizabeth Dinces. 2014. The five myths of MMN: Redefining how to use MMN in basic and clinical research. *Brain Topography* 27(4). 553–564. DOI: 10.1007/s10548-013-0326-6.
- Sussman, Elyse S. & István Winkler. 2001. Dynamic sensory updating in the auditory system. *Cognitive Brain Research* 12(3). 431–439. DOI: 10.1016/S0926-6410(01)00067-2.
- Terken, Jacques & Dik Hermes. 2000. The perception of prosodic prominence. In Merle Horne (ed.), *Prosody: Theory and experiment: Studies presented to Gösta Bruce*, 89–127. Dordrecht: Springer Netherlands. DOI: 10.1007/978-94-015-9413-4_5.
- Tiitinen, Hannu, Patrick May, Kalevi Reinikainen & Risto Näätänen. 1994. Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature* 372(6501). 90–92. DOI: 10.1038/372090a0.
- Toepel, Ulrike, Ann Pannekamp & Kai Alter. 2007. Catching the news: Processing strategies in listening to dialogs as measured by ERPs. *Behavioral and Brain Functions* 3(1). 53. DOI: 10.1186/1744-9081-3-53.
- Tsang, Yiu-Kei, Shiwei Jia, Jian Huang & Hsuan-Chih Chen. 2011. ERP correlates of pre-attentive processing of Cantonese lexical tones: The effects of pitch contour and pitch height. *Neuroscience Letters* 487(3). 268–272. DOI: 10.1016/j.neulet.2010.10.035.
- Unger, Sanford M. 1964. Habituation of the vasoconstrictive orienting reaction. *Journal of Experimental Psychology* 67(1). 11–18. DOI: 10.1037/h0044510.
- Vachon, François, Robert W. Hughes & Dylan M. Jones. 2012. Broken expectations: Violation of expectancies, not novelty, captures auditory attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 38(1). 164–177. DOI: 10.1037/a0025054.
- van Rij, Jacolien, Petra Hendriks, Hedderik van Rijn, R. Harald Baayen & Simon N. Wood. 2019. Analyzing the time course of pupillometric data. *Trends in Hearing* 23. 1–22. DOI: 10.1177/2331216519832483.
- van Rij, Jacolien, Martijn Wieling, R. Harald Baayen & Hedderik van Rijn. 2022. *Itsadug: Interpreting time series and autocorrelated data using GAMMs*. <https://cran.r-project.org/web/packages/itsadug/index.html>.
- Velden, Manfred. 1978. Some necessary revisions of the neuronal model concept of the orienting response. *Psychophysiology* 15(3). 181–185. DOI: 10.1111/j.1469-8986.1978.tb01359.x.
- Vella, Alexandra. 1995. *Prosodic structure and intonation in Maltese and its influence on Maltese English*. University of Edinburgh. (PhD thesis).

References

- Vella, Alexandra. 2007. The phonetics and phonology of wh-question intonation in Maltese. In *Proceedings of the 16th International Congress of Phonetic Sciences*, 1285–1288. Saarbrücken, Germany.
- Vella, Alexandra. 2009. Maltese intonation and focus structure. In Ray Fabri (ed.), *Maltese linguistics: A snapshot; in memory of Joseph A. Cremona (1922 - 2003)*, vol. 1, 63–92. Brockmeyer Verlag.
- Vella, Alexandra. 2011. Alignment of the ‘early’ HL sequence in Maltese falling tune wh-questions. In *Proceedings of the 17th International Congress of Phonetic Sciences*, 2062–2065. Hong Kong, China.
- Vella, Alexandra. 2013. Languages and language varieties in Malta. *International Journal of Bilingual Education and Bilingualism* 16(5). 532–552. DOI: 10.1080/13670050.2012.716812.
- Ventura, Caterina. 2020. *Production, perception and online processing of prominence in the post-focal domain*. Universität zu Köln. (PhD thesis). <http://www.uni-koeln.de/>.
- Ventura, Caterina, Martine Grice, Michelina Savino, Diana Kolev, Ingmar Brilmayer & Petra B. Schumacher. 2020. Attention allocation in a language with post-focal prominences. *NeuroReport* 31(8). 624–628. DOI: 10.1097/WNR.0000000000001453.
- von Heusinger, Klaus & Petra B. Schumacher. 2019. Discourse prominence: Definition and application. *Journal of Pragmatics* 154. 117–127. DOI: 10.1016/j.pragma.2019.07.025.
- Wagner, Michael & Duane G. Watson. 2010. Experimental and theoretical advances in prosody: A review. *Language and cognitive processes* 25(7-9). 905–945. DOI: 10.1080/01690961003589492.
- Wallace, Gerald & James M. Kauffman. 1973. *Teaching children with learning problems*. Columbus: Merrill.
- Wang, Chin-An & Douglas P. Munoz. 2015. A circuit for pupil orienting responses: Implications for cognitive modulation of pupil size. *Current Opinion in Neurobiology* 33. 134–140. DOI: 10.1016/j.conb.2015.03.018.
- Wang, Lin, Marcel Bastiaansen, Yufang Yang & Peter Hagoort. 2011. The influence of information structure on the depth of semantic processing: How focus and pitch accent determine the size of the N400 effect. *Neuropsychologia* 49(5). 813–820. DOI: 10.1016/j.neuropsychologia.2010.12.035.
- Watson, Duane G., Michael K. Tanenhaus & Christine A. Gunlogson. 2008. Interpreting pitch accents in online comprehension: H* vs. L+H*. *Cognitive Science* 32(7). 1232–1244. DOI: 10.1080/03640210802138755.
- Wechsler, David. 1987. *WMS-R: Wechsler memory scale–Revised: Manual*. San Antonio: Psychological Corporation.

- Wertheimer, Max. 1923. Untersuchungen zur Lehre von der Gestalt ii. *Psychologische Forschung* 4(1). 301–350. DOI: 10.1007/BF00410640.
- Wetzel, Nicole, David Buttellmann, Andy Schieler & Andreas Widmann. 2016. Infant and adult pupil dilation in response to unexpected sounds. *Developmental Psychobiology* 58(3). 382–392. DOI: 10.1002/dev.21377.
- Wickham, Hadley. 2016. *Ggplot2: Elegant graphics for data analysis*. Cham: Springer International Publishing. DOI: 10.1007/978-3-319-24277-4.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo & Hiroaki Yutani. 2019. Welcome to the Tidyverse. *Journal of Open Source Software* 4(43). 1686. DOI: 10.21105/joss.01686.
- Widmann, Andreas, Erich Schröger & Burkhard Maess. 2015. Digital filter design for electrophysiological data – A practical approach. *Journal of Neuroscience Methods*. Cutting-edge EEG Methods 250. 34–46. DOI: 10.1016/j.jneumeth.2014.08.002.
- Winn, Matthew B., Dorothea Wendt, Thomas Koelewijn & Stefanie E. Kuchinsky. 2018. Best practices and advice for using pupillometry to measure listening effort: An introduction for those who want to get started. *Trends in Hearing* 22. 1–22. DOI: 10.1177/2331216518800869.
- Winter, Bodo & Martijn Wieling. 2016. How to analyze linguistic change using mixed models, growth curve analysis and generalized additive modeling. *Journal of Language Evolution* 1(1). 7–18. DOI: 10.1093/jole/lzv003.
- Wolff, Susann, Matthias Schlesewsky, Masako Hirotani & Ina Bornkessel-Schlesewsky. 2008. The neural mechanisms of word order processing revisited: Electrophysiological evidence from Japanese. *Brain and Language* 107(2). 133–157. DOI: 10.1016/j.bandl.2008.06.003.
- Wood, Simon. 2017. *Generalized additive models: An introduction with R*. 2nd edn. Boca Raton: Chapman & Hall/CRC. DOI: 10.1201/9781315370279.
- Wood, Simon. 2023. *Mgcv: Mixed GAM computation vehicle with automatic smoothness estimation*. <https://cran.r-project.org/web/packages/mgcv/index.html>.
- Wright, Richard D. & Lawrence M. Ward. 2008. *Orienting of attention*. Oxford: Oxford University Press.
- Zekveld, Adriana A., Thomas Koelewijn & Sophia E. Kramer. 2018. The pupil dilation response to auditory stimuli: Current state of knowledge. *Trends in Hearing* 22. DOI: 10.1177/2331216518777174.

References

- Zellin, Martina, Andreas Pannekamp, Ulrike Toepel & Elke van der Meer. 2011. In the eye of the listener: Pupil dilation elucidates discourse processing. *International Journal of Psychophysiology* 81(2). 133–141. DOI: 10.1016/j.ijpsycho.2011.05.009.

Name index

- Alain, Claude, 22, 86, 125, 143, 145
Alamia, Andrea, 7, 34, 35, 151
Albert, Aviad, 40, 46, 64, 101
Amiez, Céline, 33
Anwyl-Irvine, Alexander L., 106
Arvaniti, Amalia, 39, 40, 42, 146
Assaneo, Florencia M., 90, 91
- Bach, Dominik R., 13, 23, 25, 48, 61,
127, 143, 145
Baddeley, Alan, 134
Balkenius, Christian, 33, 34, 151
Barr, Dale J., 72, 162, 163
Baumann, Stefan, 2, 3, 40, 41, 45–48,
50, 52, 53, 55, 83, 87–90, 96,
149
Beatty, Jackson, 35
Beaumont, J. Graham, 27
Beckman, Mary E., 42, 45, 46
Ben-Shakhar, Gershon, 14
Berti, Stefan, 132–134, 154
Bishop, Jason, 3, 55, 87, 150
Boersma, Paul, 64, 101
Bolinger, Dwight, 41, 48
Bornkessel-Schlesewsky, Ina, 27, 49,
50, 87, 92
Brandreth, Gyles, 155
Braunschweiger, David, 17
Brody, Nathan, 107
Brouwer, Harm, 50, 150
Bruggeman, Anna, 44
Byatt, Andy, 62, 68
- Bürkner, Paul-Christian, 71, 160
- Campbell, Kenneth, 3, 23, 25, 48, 61,
127, 143
Cangemi, Francesco, 2, 40, 52, 64, 101
Carrigan, Donna, 155
Chandrasekaran, Bharath, 54, 150
Chen, Aoju, 55, 96
Chen, Yiya, 51, 91, 144
Chobert, Julie, 24, 125
Clark, Andy, 53
Clifton, Rachel K., 26
Cole, Jennifer, 3, 46, 55, 62
Contier, Friederike, 151
Corbetta, Maurizio, 16
Cowan, Nelson, 14, 25
- Delorme, Arnaud, 70
Demberg, Vera, 94
Diascro, Matthew N., 107
Dimberg, Ulf, 26
Dimitrova, Diana V., 50, 150
Doeller, Christian F., 23, 86, 125, 143,
145
Donchin, Emanuel, 21, 28, 32
Duncan, Connie C., 21, 31, 32, 72
Dynes, Robin, 155
- Edwards, Jan, 46
Elliott, Emily M., 14
Emmendorfer, Alexandra K., 31, 72
Engle, Randall W., 134, 154
Eriksen, Barbara A., 105, 106

Name index

- Eriksen, Charles W., 105, 106
Escera, Carles, 25, 32, 151
Eysenck, Hans J., 105, 107
- Fothergill, Alastair, 62, 68
Fraundorf, Scott H., 49
Frazier, Lyn, 42
Frearson, Warwick, 105, 107
Friederici, Angela D., 71
Frischkorn, Gidon T., 96, 97
Friston, Karl, 15, 49, 53, 83, 130, 150
- Gandour, Jackson, 90
Gartenberg, Robert, 45, 46
Gati, Itamar, 14
Gibson, James J., 19
Gold, Joshua I., 34, 35
Grabe, Esther, 55, 96
Graham, Frances K., 26
Gregory, Richard L., 19
Grice, Martine, 2, 3, 6, 39–48, 51, 53,
54, 59, 89, 95, 102, 109, 124,
125, 129, 146, 149, 150
Guo, Zhe-chen, 51, 52
Gussenhoven, Carlos, 3, 42, 45, 46
- Harris, Jesse A., 93
Hart, John T., 45
Hasting, Anna S., 33
Hatfield, Gary, 16
Heitz, Richard P., 134, 154
Hermes, Dik, 2, 40
Hickok, Gregory, 90
Himmelmann, Nikolaus P., 2, 40, 52
Hoeks, John C. J., 50, 150
Hommel, Bernhard, 11
Horowitz, Frances Degen, 49
Hsu, Chun-Hsien, 24, 49, 52, 85, 86,
125, 143, 145
- Hsu, Yi-Fang, 32, 151
Huettig, Falk, 53
Hughes, Robert W., 16, 25, 126, 133,
154
Häuser, Katja I., 94
- James, William, 1, 11, 12, 16, 17
Jesse, Harris A., 94
Johansson, Birger, 33, 34, 151
Johnson, Elizabeth K., 52
Jones, Dylan M., 25
Joshi, Siddhartha, 34, 35
Jun, Sun-Ah, 6, 42–44, 46, 51, 94, 156
- Kakouros, Sofoklis, 3, 50, 55, 87, 150
Kauffman, James M., 155
Kember, Heather, 51
Keye, Doris, 97, 134
Knight, Rachael-Anne, 3, 41, 45, 46
Koch, Xaver, 49
Koffka, Kurt, 17
Kohler, Klaus, 45, 46
Kreegipuu, Kairi, 33
Kyröläinen, Aki-Juhani, 110
Kügler, Frank, 2, 40, 51
- Ladd, D. Robert, 3, 39, 41–43, 45, 46,
51, 60, 61, 95, 129, 146
Lauchlan, Fraser, 155
Leiner, Dominik J., 104, 160
Li, Weijun, 50, 87
Li, Xiaoqing, 51, 91, 144
Lialiou, Maria, 44, 59, 93, 125, 156
Liao, Hsin-I, 3, 16, 35, 94, 125, 143
Lieberman, Mark Yoffe, 41, 45
Liu, Min, 51, 53, 84, 87, 91, 144, 149
Lorenzen, Janne, 41, 46, 48
Luck, Steven J., 27–29, 31, 33
Lynn, Richard, 15

- Macdonald, Margaret, 3, 23, 25, 48,
61, 127, 143
Maess, Burkhard, 71
Makeig, Scott, 70
Maltzman, Irving, 16–18, 26
Mandell, Mary P., 17
Marois, Alexandre, 16, 34, 35, 95, 125,
143, 151
Mathôt, Sebastiaan, 104, 160
Milner, Peter M., 19
Morton, Rachel, 3, 45, 46
Munoz, Douglas P., 34, 151
Mücke, Doris, 46
Müller, Hermann J., 19

Niebuhr, Oliver, 3, 41, 45, 46
Nieuwenhuis, Sander, 34, 35, 151
Näätänen, Risto, 1, 3, 7, 11, 12, 14–16,
18, 20–22, 25–27, 29–33, 50,
59, 61, 68, 70, 71, 83, 86, 91,
125, 130, 142, 143, 145, 147,
151

Oldfield, Richard C., 62
Orrico, Riccardo, 156
Ou, Shu-chen, 51, 52

Paavilainen, Petri, 15, 30, 126
Pavlov, Ivan Petrovich, 1, 13, 14, 16,
18, 141
Peters, Jörg, 55, 56, 83, 96
Petersen, Steven E., 11, 12, 34
Pierrehumbert, Janet B., 41, 42, 45
Poeppel, David, 90
Polich, John, 21, 31, 32, 144
Posner, Michael I., 1, 11, 12, 14, 16, 19,
34
Primus, Beatrice, 2, 40, 52
Procyk, Emmanuel, 33

Quasthoff, Uwe, 62

Rabbitt, Patrick M., 19
Raskin, David C., 17
Rathcke, Tamara, 45
Reisberg, Daniel, 14
Ren, Guiqin, 51, 87, 91, 144
Repp, Magdalena, 49, 150
Richter, Matthias, 62
Rietveld, Toni, 3, 45, 46
Rimland, Kim, 55, 87, 150
Rinne, Teemu, 3, 22, 23, 25, 127, 143
Roessig, Simon, 41, 46, 53
Roettger, Timo B., 55, 87, 150
Rosenberg, Andrew, 45
Ruchkin, Daniel S., 50
Röhr, Christine T., 3, 33, 41, 46, 47,
49–51, 53, 83, 84, 86–91, 95,
109, 124, 125, 128, 129, 144,
146, 149, 150

SanMiguel, Iria, 97, 132–134, 154
Savino, Michelina, 3, 49, 51, 95, 109,
124, 129, 146
Sayeed, Asad, 94
Scheepers, Christoph, 94
Schlesewsky, Matthias, 50
Schröger, Erich, 132–134, 154
Schumacher, Petra B., 2, 27, 40, 41,
48–50, 83, 150
Seidl, Amanda, 52
Shattuck-Hufnagel, Stefanie, 42
Shulman, Gordon L., 16
Smith, Rachel, 45
Sokolov, Evgeniï N., 1, 12, 14, 145
Sokolov, Evgeniï Nikolaevich, 1, 12,
14, 87, 145
Spalek, Katharina, 49
Spilsbury, Georgina, 111

Name index

- Strauch, Christoph, 8, 33–35, 94, 126,
144, 151
Streefkerk, Barbartje M., 2, 40, 52
Sullivan, Joseph W., 49
Sussman, Elyse S., 15, 31, 83, 147
Sóskuthy, Márton, 114
Sörqvist, Patrik, 97, 133, 134, 154

Terken, Jacques, 2, 40
Tiitinen, Hannu, 16
Toepel, Ulrike, 50, 150
Trouvain, Jürgen, 55, 83, 96
Tsang, Yiu-Kei, 51, 87, 91, 144, 146
Turk, Alice E., 42

Unger, Sanford M., 15, 26

Vachon, François, 15, 83, 125, 126, 147
Van Rij, Jacolien, 110, 114
Velden, Manfred, 15
Vella, Alexandra, 44, 156
Ventura, Caterina, 3, 49, 54, 87, 125,
146, 150
Von Heusinger, Klaus, 2, 40

Wagner, Michael, 42
Wallace, Gerald, 155
Wang, Chin-An, 34, 151
Wang, Lin, 49
Ward, Lawrence M., 16, 17, 20, 126
Watson, Duane G., 42
Wechsler, David, 105, 109
Weenink, David, 64, 101
Wertheimer, Max, 17
Wetzel, Nicole, 3, 35, 95, 143
Wickham, Hadley, 71, 110, 160
Widmann, Andreas, 71
Wieling, Martijn, 114
Winkler, István, 15, 20, 83, 147
Winn, Matthew B., 34, 94

Winter, Bodo, 2, 3, 40, 41, 45–47, 90,
114
Wolff, Susann, 71
Wood, Simon, 110, 114
Wright, Richard D., 16, 17, 20, 126

Zekveld, Adriana A., 34
Zellin, Martina, 94

Prosody and attention orienting

Humans are confronted everyday with an influx of sounds coming from several sources. In a given auditory environment, some of the sound events might be unexpected, rare, or new. Our cognitive system has the ability to detect such sounds, and consequently activate an attention orienting response.

This book provides an in-depth investigation of the interplay between prosody and attention orienting during online speech processing by using two complementary experimental methods, electrophysiology and pupillometry. In particular, it examines the cognitive and functional relevance of intonation for the orienting response in speech, emphasising the crucial role of rising contours. More specifically, it investigates the influence of prosodic structure, showing that rising tones at constituent edges attract attention similarly to accentual rising tones, challenging important tenets of prosodic theory and typology.

The book shows that contextual expectations shape the orienting response, extending insights beyond purely acoustic cues to pragmatically meaningful linguistic signals. Finally, much like cues from auditory cognition, rising tones activate both involuntary and voluntary attentional mechanisms – the former driven by signal-based acoustic cues, while the latter arise from contextual influences guiding voluntary attention orienting.

Taken together, the findings in this book enhance our understanding of the role of intonation in attention orienting, emphasising the significance of rising tones. The book establishes key connections to general cognition and individual variability while exploring potential extensions for an architecture of attention orienting in spoken language.