



Individual variability in the attention-orienting function of edge tones in German: A pupillometry study

Maria Lialiou*, Department of German Language and Literature I, Linguistics, University of Cologne; IfL Phonetik, University of Cologne, mlialiou@uni-koeln.de

Jesse A. Harris, Department of Linguistics, UCLA

T. Mark Ellison, IfL Phonetik, University of Cologne

Petra B. Schumacher, Department of German Language and Literature I, Linguistics, University of Cologne

Martine Grice, IfL Phonetik, University of Cologne

*Corresponding author.

Using a *changing-state* oddball paradigm, in which auditory sequences of sequentially ordered ascending numbers (*standards*; e.g., 21 22 23 24 25 26...) are occasionally interspersed with an out-of-sequence number (*deviant*; e.g., 25 in 21 22 23 25 26...), we investigate the role of individual variability in the effect of rising intonation on attention orienting in German. Standard numbers featured a shallow fall in pitch, referred to as neutral intonation, as it is a pattern commonly found on nonfinal items in a sequence, while deviant numbers were produced with either the same neutral intonation, or rising or falling intonation. The difference in pitch contour was a reflex of different edge tone combinations, the pitch accent being the same in all three conditions. Participants' cognitive variability was measured based on processing speed, inhibitory ability, and working memory capacity (WMC), using odd-man-out, flanker, and digit span tasks, respectively. Results show that individuals with better inhibitory skills, showed sustained pupil dilation response (PDRs) to rising deviants, compared with neutral and falling ones, with no difference between the latter two ([rise] > [fall = neutral]). In contrast, individuals with weaker inhibitory skills exhibited sustained PDRs to both rising and falling intonation, but not to the baseline neutral ([rise = fall] > [neutral]). These findings highlight individual differences in the allocation of attentional resources and suggest that phrase-final pitch movements effectively capture attention across cognitive profiles. This supports the view that intonational tones at the edges of constituents can serve as attention-orienting devices, with implications for theories of intonation and auditory attention.



1. Introduction

A fundamental characteristic of the human cognitive system is its tendency to withdraw attention from the current locus of attention and orient it towards a source of information that may be 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 attentional resources towards the deviant auditory event (for reviews, see Näätänen, Kujala, & Light, 2019). Attention orienting has been investigated primarily in cognitive domains like vision and audition. In particular, it has been shown that auditory attention orienting is sensitive to the saliency of the physical properties of the signal. It has been found that rises in the amplitude or pitch of sine waves activate attentional resources, which in turn elicit a series of enhanced neural, psychological, or physiological reflexive responses (e.g., Näätänen et al., 1978; Rinne et al., 2005, 2006; Bach, 2008; Macdonald & Campbell, 2011, among others).

We focus here on the role of intonation in attention orienting. Intonation is traditionally defined as “the use of *suprasegmental* phonetic features to convey ‘postlexical’ or *sentence-level* pragmatic meaning in a *linguistically structured way*” (Ladd, 2008: 4). Intonation is used to signal, among other things, highlighting (linked to a prominence-cueing function) and phrasing. Highlighting and phrasing are expressed through phonological choices—such as pitch accents (tonal movements tied to stressed syllables) and edge tones (tonal movements marking the boundaries of prosodic units)—as well as through acoustic parameters (e.g., scaling and timing of fundamental frequency (f_0), intensity, segmental durations, and spectral features). In the present study, we concentrate specifically on the role of edge tones in orienting attention. A recent study by Lialiou et al. (2024) explored whether the attentional system is differentially oriented towards rising tones (both as pitch accents and edge tones). Specifically, the role of rising intonation for attention orienting was investigated using event-related potentials (ERPs). The results of this study showed that 1) intonational rises are especially effective in reorienting attention (in contrast to intonational falls), and 2) the attention-orienting function of rising intonation in speech holds regardless of whether the rise was accentual or associated with an edge tone. This result is not predicted by standard intonational theory, as edge tones are regarded as serving a segmentation function (phrasing) and are therefore excluded from contributing to prominence cueing (as discussed in Grice, 2022). Capitalising on these findings, the current study explores the attention-orienting function of domain-final intonational rises and falls (which are the reflex of phrase-final edge tones) in German. In the remainder of the paper, we use the terms *domain-final*, *edge*, and *boundary* interchangeably when indicating tonal position.

In addition to investigating general patterns, the present study also examines how individual variability in attentional control influences the processing of such intonational cues. Individuals vary in the way they use or control attentional resources towards incoming information. 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-sequence number (*deviant*), we investigate the role of individual variability in the effect of edge tones on attention orienting, by measuring listeners' *pupil dilation response* (PDR). We further seek to link the differences in PDR to cognitive measures of individual variation.

1.1. Attention orienting and pupillary responses

The concept of *attention orienting* is rooted in the seminal work of Ivan Pavlov. Pavlov (1927) described the concept of an *orienting response* (OR) 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. Pavlov proposed an association between the initiation of the OR and an early processing stage. During this processing stage, the organism is prepared 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 refined Pavlov's OR concept and have proposed various mechanisms to explain it. One such mechanism 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 OR (for the mechanisms of attention orienting, see Näätänen, 1990; Sussman & Winkler, 2001; Friston, 2010, 2018; Näätänen et al., 2011; Vachon et al., 2012, among others).

It has been shown that the OR finds expression in behavioural factors (e.g., Näätänen et al., 1993; Cowan, 1998; Escera et al., 1998; Näätänen et al., 2001; Hughes & Jones, 2003, 2005) as well as multiple physiological indices, such as electrodermal (e.g., Maltzman, 1979), electromyographic (e.g., Dimberg, 1990), vascular (e.g., Unger, 1964), cardiac (e.g., Graham & Clifton, 1966), neural (for review of the different brain measurements, see Näätänen, 1992), and pupillary measures (e.g., Nieuwenhuis et al., 2011; Wang & Munoz, 2015; Johansson & Balkenius, 2018; Marois et al., 2018; Alamia et al., 2019). As mentioned in Johansson & Balkenius (2018), it has long been established that although large changes in pupil size emerge as a result of luminance changes, pupillary responses also occur as a function of distinct brain processes. The PDR has 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 deviances 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).

Pupillometry provides a relatively high resolution of temporal information, making it possible to estimate cognitive activity related to attentional processes over time. Attention-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). Pupillometric studies on auditory attention, using oddball paradigms, have further reported that attention-related PDRs are not only induced by the presence of an auditory deviant, but are also sensitive to the saliency of the physical characteristics of that deviant sound. 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).

High-level factors can also impact attention-related PDRs (e.g., Joshi & Gold, 2020; Strauch et al., 2022). This is reflected in the PDR manifesting a more sustained response (for more, see Strauch et al., 2022). Accordingly, the latency of attention-related PDRs is indicative of the awareness level or processing stage: whereas transient responses occur preattentively, more prolonged responses emerge in full awareness (see Strauch et al., 2022). The route from preattentive to conscious perception of a deviant is therefore reflected in the time course of pupil responsiveness (in the words of Strauch et al., 2022, an integrated readout of attentional networks). One could hence argue that transient, signal-sensitive PDR is an instantiation of an 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 align with neurophysiological and linguistic research on auditory and speech perception, showing that signal-based properties are essential for attracting involuntary attention, while voluntary attention is further activated by more complex top-down processes. In auditory attention, the saliency of an event is essential: the greater the rise in amplitude or pitch of a deviant sound, the greater the orienting response (e.g., Näätänen et al., 1978; Rinne et al., 2005, 2006; Macdonald & Campbell, 2011). Similarly, in speech perception, intonational rises are used for attracting interlocutors' attention when asking questions (e.g., Dingemans et al., 2013), directing listeners' attention towards the most important part and/or an unexpected change in an utterance (e.g., Röhr et al., 2021; Lialiou et al., 2024), and even guiding attention in serial recall tasks (e.g., Savino et al., 2020 for edge tones, Grice et al., 2024 for both pitch accents and edge tones).

1.2. Intonational prominence, edge tones, and attention

Spoken communication necessarily involves *intonation*, the melody or the so-called tune of an utterance. It is encoded primarily through modulations in fundamental frequency (f_0 ; e.g., Liberman, 1975; Pierrehumbert, 1980; Ladd, 2008; Beckman, 2012), perceived as pitch variation, but other acoustic properties such as amplitude and intensity (perceived as loudness) and duration (perceived as length), among others, also play a role.

Intonation serves multiple functions, from helping to locate fixed attributes of words—such as lexical stress in West Germanic languages—to marking attributes at the utterance level (for a comprehensive review, see Grice et al., 2023). Prominence is defined as the property of a linguistic unit “standing out” from its neighbouring environment. In intonation research, prominence pertains to the form that is used to indicate this property of standing out (e.g., Terken & Hermes, 2000; Streefkerk, 2002; Cangemi & Baumann, 2020; Grice & Kügler, 2021). Whilst speakers may use a wide range of linguistic means for signalling prominence at the utterance level, including prosodic and nonprosodic factors (e.g., Baumann & Winter, 2018), intonation is a pivotal cue to prominence in spoken language. Specifically, speakers can make use of intonation to highlight important parts of their message, thereby orienting listeners’ attention to specific information (Ladd, 2008, ch. 6; Chafe, 1974). At the same time, 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 at privileged positions in the phrase, referred to as an *intonational rise*.

In West Germanic languages, modulations in f₀ direction, excursion, scaling, and timing are some of the acoustic dimensions found to be indicative of different degrees of perceived prominence, such that f₀ peaks are generally perceived as more prominent than f₀ 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 f₀ 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 f₀ contour (e.g., the steepness of a rise or fall as well as its alignment with a stressed syllable) is pivotal to prominence (e.g., Kohler & Gartenberg, 1991; Niebuhr, 2009; Knight, 2008).

In German, prominence is encoded (and decoded) through the use of pitch accents on stressed syllables. The identity of a pitch accent is defined by its phonetic substance on the basis of the f₀ dimensions direction (rise/fall), scaling (steep/shallow), height (peak/valley), and alignment (timing of f₀ 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.¹ 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) revealed the importance of intonational rises, as higher f₀ peaks and steeper f₀ rises are perceived as more prominent.

¹ Pitch 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 paper, and will therefore not be discussed further.

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). Theories of intonational phonology, especially the autosegmental-metrical model (AM; e.g., Ladd, 2008; Arvaniti & Ladd, 2023; Arvaniti, Grice & D’Imperio, 2025), and prosodic typology (e.g., Jun, 2014) postulate that tonal events come prepackaged with specific associations and functional properties (in the case of English, German, and Italian; see Grice, 2022 for discussion). Importantly, in head prominence languages, pitch accents are assumed to take on the role of highlighting selected constituents, while edge tones merely serve to chunk utterances into units. These functions relate to Jun’s prosodic typology, where German, English, and Italian are head prominence languages as opposed to, e.g., Korean, an edge prominence language. Nonetheless, two serial recall studies (e.g., Savino et al., 2020 on Italian, Grice et al., 2024 on German) have also provided some evidence for edge prominence in canonically head prominence languages. Specifically, Grice et al. (2024) report that rising edge tones in German attract attention by highlighting the whole domain they delimit, as reflected in listener recall performance. This highlighting of the domain is not merely a result of a grouping effect (e.g., Sturges & Martin, 1974; Reeves, Schmauder & Morris, 2000). Crucially, a rise has a greater effect on recall than a fall, all other things being equal—in this case the strength of the boundary (Intonational Phrase, hereafter IP). Further, a recent ERP study in German has shown that rises associated with the edges of constituents induce an attention-orienting response in a similar way to accentual rises (Lialiou et al., 2024). Last but not least, processing studies on the domain of the word (see Grice & Kügler, 2021) reported that word segmentation (e.g., Ou & Guo, 2021) and word recognition (e.g., Kember et al., 2021) is improved when boundary rises mark those words. Therefore, although it is claimed by AM that in West Germanic languages prominence is signalled through pitch accents, directing listener attention towards key information, it appears that prosodic boundaries also play a role in making a word or a larger constituent more prominent. This use of distinctive intonation at prosodic boundaries might be due to the necessity of highlighting key information in critical positions for speech processing and planning, such as at the beginning or end of an utterance (e.g., Seidl & Johnson, 2006; Ou & Guo, 2021).

1.3. Individual variability in speech and auditory attention

Research has traditionally prioritised group-level patterns over individual differences, aiming to uncover generalisable trends across populations. This focus often stems from the desire for methodological simplicity and statistical robustness. In recent years, individual variability has increasingly attracted attention in both linguistic and (neuro)cognitive studies, as researchers have recognised that group averages may obscure meaningful within-group differences.

Zooming in first on speech, studies on the prosodic marking of prominence have shown that speakers encode prominence relations using prosodic cues in different combinations and with different degrees of strength (e.g., Baumann & Winter, 2018; Lorenzen et al., 2024). In

a similar fashion, production studies on speech acts report variability in the prosodic means that individuals use to encode exclamative utterances and rejecting questions (e.g., Repp, 2020; Repp & Seeliger, 2020; Seeliger and Repp, 2023). Further, work on the perception of prosody suggests that individuals differ in this domain, too. Such studies have explored the role of social cognition in the perception of prosody, and showed that individuals with higher communicative or pragmatic skills are more sensitive to the decoding of prominence relations, information structure and prosodic structure (e.g., Bishop, 2012, 2016; Jun & Bishop, 2015a/b; Hurley & Bishop, 2016; Bishop et al, 2020).

Individuals can also vary in how they use or control attentional resources in processing incoming information. These differences might arise from the use and activation of different cognitive functions interacting with the attentional system. Studies that are designed to better understand individual variability during cognitive processing usually examine cognitive functions like working memory (WM) capacity, processing speed, and executive processes (e.g., inhibition, shifting, updating), among others (for review, see Frischkorn et al., 2022). However, previous research on individual variation in auditory attention has yielded sparse and contradictory evidence. Starting with the role of WM capacity, whereas some studies have reported that individuals with large WM capacity are less susceptible to attentional switches towards auditory deviations (for review, see Sörqvist et al., 2013; Hughes, 2014), other studies have shown that increased WM load (or low WM capacity) attenuates or even prevents auditory distraction (e.g., Berti & Schröger, 2003; SanMiguel et al., 2008). In terms of prosodic discrimination, Stepanov et al. (2020) found that children with better WM for both storing and processing sounds performed more accurately in prosodic discrimination, although having high capacity in both components did not confer additional benefits. Nonetheless, it is important to note that what happens during acquisition may be different from adult processing. Moreover, in an imitation task, Petrone et al. (2021) showed that higher WM facilitated phonological imitation, particularly for prosodically rich read speech. However, WM did not have an effect on the more variable phonetic imitation. Similarly, van der Burght et al. (2025) reported that listeners with higher WM spans classified prosodic structures more accurately and required less information to detect boundaries. This effect was independent of processing speed, attention, or motivation, suggesting that individual differences in WM partly account for variability in prosody perception. As mentioned in Sörqvist et al. (2013), the exact nature of the mechanism that WM taps into is still under debate, with one of the views claiming that individuals with large WM capacity have excellent inhibitory skills. It is important to note that in the relevant studies, a link between WM and inhibition was presumed, but not tested directly.

Moving to inhibition and processing speed, no previous study has directly tested the effect of those two functions on auditory attention. A 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) was concerned with *visual* selective attention. Keye et al. (2009) found

no support for a relation between WM and attention control or inhibition (conflict reduction). However, the authors found a relation between large WM capacity and slow speed, although, Heitz & Engle (2007), another study again on visual attention, reported the converse, i.e., that individuals with large WM capacity had faster responses than individuals with small WM capacity.

Overall, these findings highlight the complex and sometimes seemingly contradictory nature of individual cognitive variability. While working memory capacity appears to influence susceptibility to auditory distraction, the mechanism behind this effect (and the duration of the effect) remains unclear, particularly given the presumed but untested role of inhibition. Furthermore, the potential contributions of processing speed and inhibitory control to auditory attention are largely unexplored, and existing evidence from visual attention studies is inconsistent. Together, these results suggest that cognitive variability arises from the interaction of multiple functions, and that further research is needed to clarify how these factors jointly shape attentional control.

1.4. Motivation for the current study

The current study examines how attention orienting in response to deviances in numeric sequences is modulated by the intonation pattern that these deviances feature. This study builds on the serial recall studies by Savino et al. (2020) and Grice et al. (2024), on Italian and German respectively, mentioned in Section 1.2, as well as the recent finding by Lialiou et al. (2024) on German that rises associated with the edges of constituents induce an attention-orienting response similar to accentual rises, the present study focuses on the processing of rising and falling edge tones (in German), investigating the effect of edge tones on attention orienting. Further, departing from studies focusing on group-level effects, we take an individual-differences approach, exploring the role of individual cognitive variability on the effect of the status and tonal properties of edge tones in attention orienting.

Listeners' pupil size, and specifically, PDR is used as a proxy for attention orienting—a linking hypothesis which has been commonly assumed in prior research. We employed an auditory changing-state oddball paradigm in passive recording of the pupil (e.g., Näätänen et al., 2019). The paradigm consisted of 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 (in red)*; e.g., 25 in 21 22 23 25 26 27 28...). Numeric sequences were selected for this study because 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 upcoming numbers. The sequences thus provide an optimal context for studying attention orienting.

Standard numbers were recorded with shallow falling intonation (hereafter, *neutral intonation*), a pattern that can also be found on nonfinal items of a list in German, while deviant numbers were produced with one of 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 tones can mark the end of either small or large units in a sequence, and they thus fulfil a similar chunking function. Therefore, the naturalistic pitch contours used in this study for the realisation of both standard and deviant numbers simulate a natural linguistic context in the form of a list.

Given that attention orienting is indexed, at least in part, by dilations in pupil size (e.g., Liao et al., 2016; Marois et al., 2018, 2019), our 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 (e.g., Hughes et al., 2007; Vachon et al., 2012; Paavilainen, 2013; Hughes, 2014; Näätänen et al., 2019). The disruption of the anticipated pattern caused by the deviant number will thus capture attention, which in turn will induce an increased PDR. Based on the reported enhanced orienting function of rising pitch compared to falling pitch (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 et al., 2024), we 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.

As a step towards better understanding the role of individual differences in attention-orienting, we employed a set of individual difference measures, to explore whether individuals with different processing capacities were affected by a numerical deviance in unique ways depending on the prosodic realisation of the deviant. We specifically focused on selective attention as measured by inhibitory ability, processing speed, and WM, resulting from the individuals' WM capacity. In what follows, we elaborate on these three cognitive abilities.

Inhibition may be one of the most crucial cognitive operations for understanding how the mechanism of attention orienting towards deviant events might vary 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 (for discussion see Frischkorn et al., 2022) is essential in explaining varying processing patterns among individuals. Therefore, processing speed might also contribute to the orienting response in that processing speed may aid or impede the evaluation of deviants, for example, on the basis of signal-driven cues like different intonation patterns (as in this study).

Lastly, WM is one of the most frequently investigated cognitive measures of individual variability. While WM has been found to be a strong factor in individual variability in general 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 WM load and mitigation of distractions at the later stages of orienting: the greater the WM load, the less susceptible a person is towards distractions (e.g., SanMiguel et al., 2008). Thus, in the context of the current study, WM measures might shed more light on the link between attention-related mechanisms and WM resistance operations. In the present study, WM *resistance* refers to the ability to maintain task-relevant information while resisting interference from irrelevant input, consistent with evidence that increased working memory load reduces susceptibility to auditory distraction (Berti & Schröger, 2003; SanMiguel et al., 2008).

2. Methods

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.

2.2. Speech materials

The auditory stimuli are 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 1** illustrates instantiations of filler and experimental materials.

The experimental items introduced arithmetic deviances, i.e., a number out of sequence, referred to as 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

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 1: Example filler and experimental materials.

increase the difficulty of the task. To control for potential effects of the position in the sequences of the deviant, it 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 1**. By varying the position of the deviant, the sequence length also varied, which served to make the deviant's position 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* ['fʏnfʦɪç]), four (e.g., 52 *zweiundfünfzig* ['tʃvaunʦfʏnfʦɪç]) or five syllables (e.g., 57 *siebenundfünfzig* ['zi:bənʦfʏnfʦɪç]), 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 given in **Figure 2**. In the rising condition, all standards in the sequence were produced with the same intonation, that is, with a shallow falling contour with an intermediate phrase edge tone, L- (hereafter, *neutral intonation*). This contour is typically used on nonfinal items of a sequence or list in German. The deviant was realised with a boundary rising intonational contour, analysed as an IP edge tone H%. In the falling condition, the standards were produced with the same neutral intonation as in the rising condition, but the deviant was realised with a boundary falling

intonational contour, an IP edge tone L%. A neutral condition, in which both standard and deviants in the sequence were produced with the shallow falling intonation, served as the baseline condition. 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.

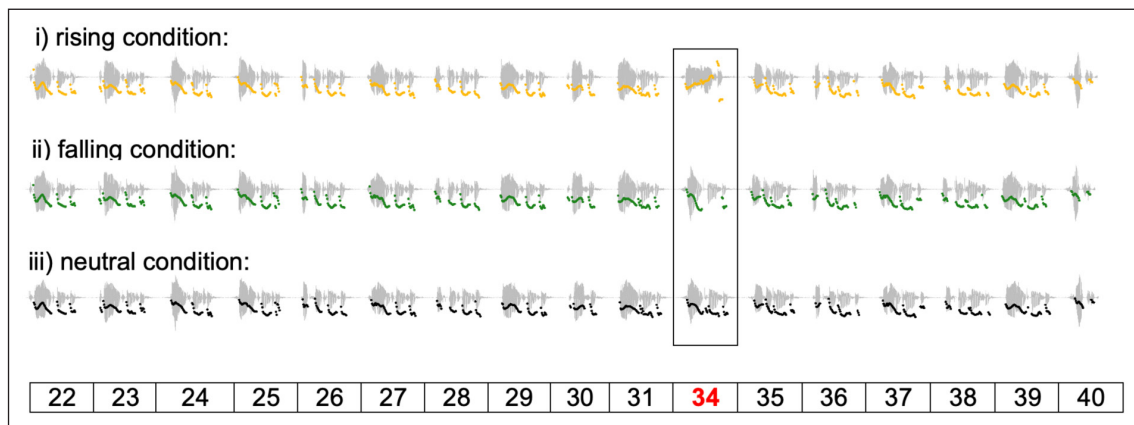


Figure 2: Speech waveform & f0 contour of a sample experimental trial sequence per condition.

The filler items were constructed without a deviant in the sequence, enhancing the expectancy creation 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/j8295/overview>). 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 and Weenink, 2024). The inter-stimulus silent interval between spliced numbers was 100 ms. The average duration of the medium-length and long sequences was 24.86 s and 25.21 s, respectively. All stimuli used in the experiment were normalised at -23 LUFS but not manipulated further.

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). Prior to the *ProPer* analysis, f_0 contours were extracted and corrected manually in Praat (Boersma and 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. 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 the first syllable, Δf_0 is calculated relative to the speaker's f_0 median. The raw Δf_0 is measured in Hz; in this analysis relative Δf_0 values are used (relative $\Delta f_0 = \text{raw } \Delta f_0 / \text{speaker's } f_0 \text{ range}$; values at zero indicate level pitch, positive values indicate high pitch, while values negative values indicate low pitch). For more on Δf_0 , see Albert (2023). **Table 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 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 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. **Figure 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

Measurement		Rising	Falling	Neutral
relative Δf_0	quadrisyllabic			
	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	pentasyllabic			
	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)

Table 1: Mean and standard deviation values (in brackets) for relative Δf_0 values (low < 0 < high) per syllable across items for each condition as well as for the total duration of items.

H* ^H-%. In both falling and neutral conditions (see **Figure 3**, depictions 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 (as suggested by the mean F0 values in **Table 1**, though not statistically tested), 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 neutral than in the falling condition.² In GToBI, the neutral and the falling conditions are labelled as H* L- and H* L-%, respectively (see **Figure 4**, depictions in green and black, respectively). The fact that the rising and falling contours involve an IP boundary, while the neutral one involves an intermediate phrase (ip) is further supported by the duration measurements in **Table 1**, which show that the final syllable exhibits greater preboundary lengthening in the rising and falling conditions, but not in the neutral condition.

² The figure shows mean Δf_0 values connected by lines and should not be read as a continuous contour. In the falling condition, the drop from the third to the fourth syllable (-15.8) is smaller/less steep than from the second to the third (-28.2), which may give the false impression that the final syllable is higher than the preceding one, although it is not.

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 smaller unit (ip), and thus can be featured on nonfinal items in a longer sequence. The falling contour, in contrast, unambiguously marks the end of a larger unit (IP). Finality is indicated by 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).

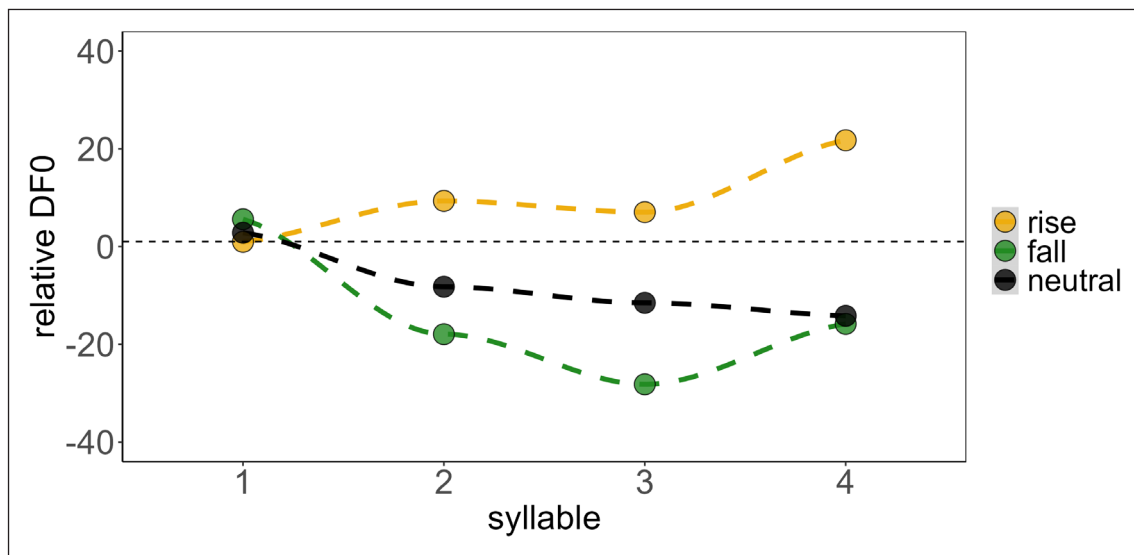


Figure 3: Mean relative Δf_0 values per syllable (x-axis) across prosodic conditions in quadrisyllabic deviant numbers.

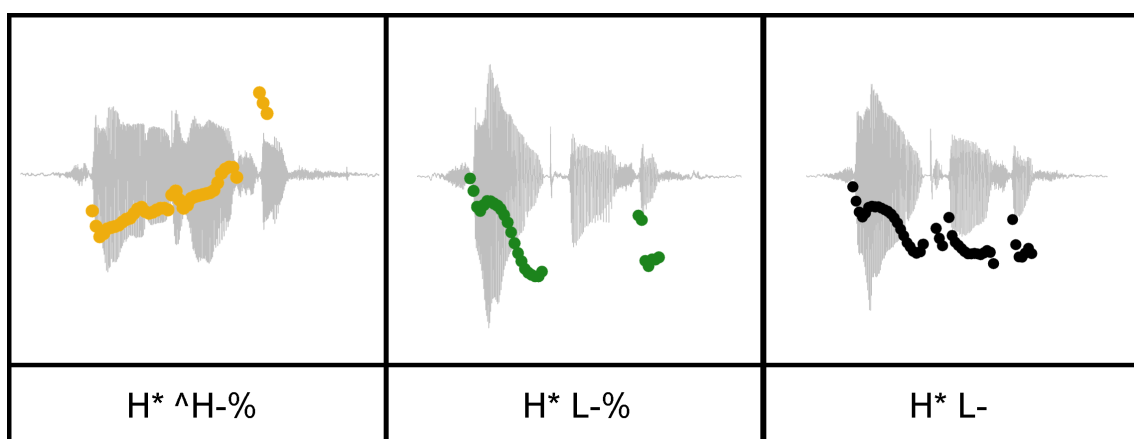


Figure 4: Instantiations of deviant f_0 contours per prosodic condition in GToBI; L-% in GToBI is equivalent to L-L%, and ^H-% is equivalent to ^H-H% in MAE ToBI.

2.3. Experimental procedures

Each experimental session consisted of the pupillometry, followed by a battery of cognitive tests (Lialiou et al., in press), 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 WM capacity, which we translate to WM resistance). 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.

2.3.1. Pupillometry

For the pupillometry, 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 the task, participants had to answer a yes/no comprehension question related to the numeric sequence they heard in 35% of the trials ($n = 25$),³ 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 black cross on the screen. 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** depicts a schematic illustration of an experimental trial.

³ For example, one of the questions was *War die Aufzählung in Zehnerschritten?* (Was the enumeration in steps of ten?) The full set of the questions are 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% and 100%), indicating high engagement.

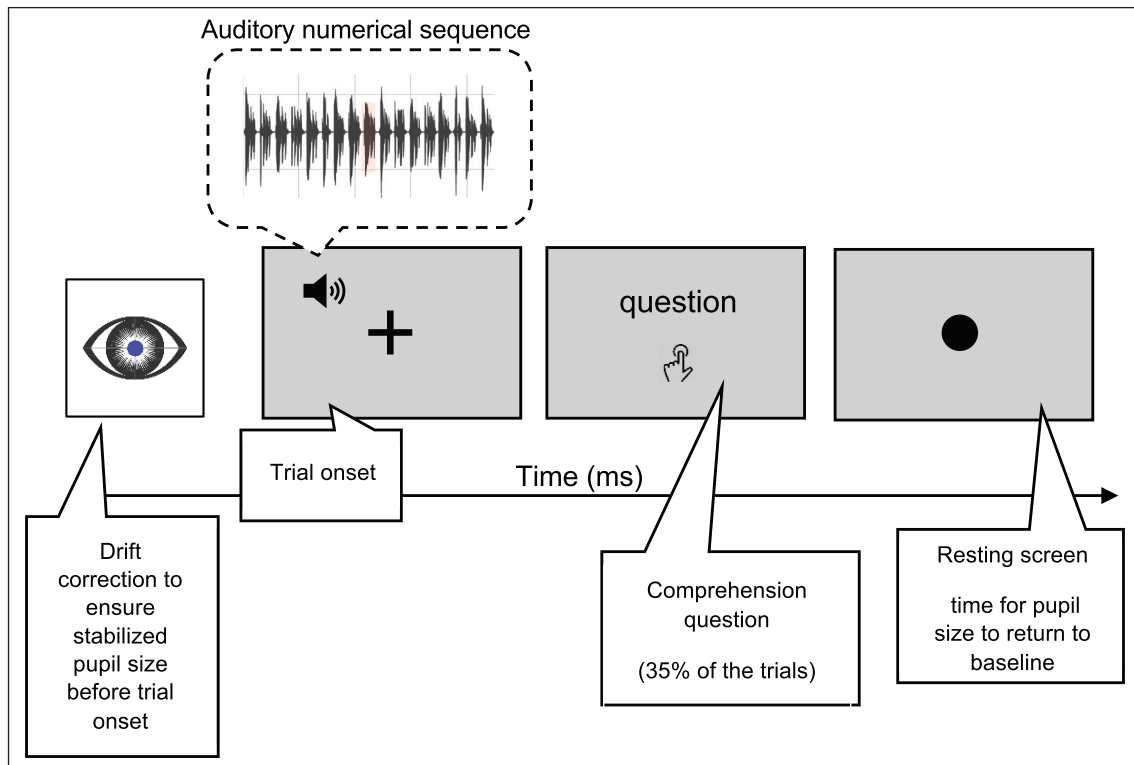


Figure 5: Schematic illustration of an experimental trial in the pupillometry.

2.3.2. 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 has been made freely available by Lialiou et al. (2025, in press) and 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 WM resistance, respectively (repository link: <https://osf.io/muh9t/overview>). Minimal descriptions of the cognitive tasks are provided below. Comprehensive details, including full task scripts and implementation parameters, are available in Lialiou et al. (in press) and in the accompanying OSF repository.

Inhibitory control was assessed with a modified arrow-based flanker task, in which participants indicated the direction of a central arrow flanked by congruent or incongruent arrows. The task comprised 96 experimental trials (48 congruent, 48 incongruent) presented in randomized blocks, with responses made via keyboard. Each trial began with a fixation cross, followed by the stimulus display until response; the task lasted approximately five minutes. Processing speed was measured with a modified odd-man-out task, in which participants quickly identified which of three hexagonal stimuli differed from the others using keyboard responses.

Each trial began with a 2000 ms fixation cross, followed by a 100 ms stimulus display and a blank screen until response, with a practice phase and 120 randomized experimental trials; the task lasted approximately five minutes. Working memory was assessed with a modified digit span task, in which participants listened to auditory digit sequences of increasing length (3–9 digits) and recalled them in order using a numeric keypad. Each trial began with an 890 ms beep and 500 ms silence, followed by the digit sequence, after which participants entered their response. The task included 14 sequences and lasted approximately five minutes.

2.4. 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 *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/j8295/overview>.

2.4.1. Data preprocessing

2.4.1.1. Pupillometry

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 artefacts were removed by hand. Subsequently, trials including more than 20% of missing data because of blink artefacts were completely removed from further analyses, yielding 3.52% loss of the total dataset. After artefact 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 artefacts 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 downsampled to a rate of 100 Hz (10 ms time bins).

2.4.1.2 Cognitive test battery

For both 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, we employed the *efficiency measure* by Spilsbury et al. (1990). Specifically, we divided the number of the correct incongruent trials by the median response time (inhibition score = [number of correct answers in incongruent trials]/[median response time]) per participant. To measure processing speed, we calculated a similar *efficiency measure* 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. The smaller the span score, the smaller the WM capacity, and thus the higher the WM resistance. In contrast, the larger the span score, the larger the WM capacity, and thus the lower the WM resistance.

The individual cognitive profiles in terms of flanker, odd-man-out, and digit span scores are presented in **Table A1** in the appendix. 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 digit span indicate smaller WM resistance; and higher scores in the odd-man-out indicate slower processing speed. The next processing step consisted of standardising (z-scoring) participants' scores across the flanker, odd-man-out, and digit span tasks. Prior to statistical modelling, correlation tests among the three cognitive scores were employed. The correlation tests revealed a weak positive correlation between odd-man-out and flanker scores ($r(58) = 0.34, p < .0001$), 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 ($r(58) = -0.23, p < .0001$), indicating that the greater the span, the lower the inhibitory skills. Finally, no correlation was found between odd-man-out and digit span scores ($r(58) = -0.06, p = .66$).

2.4.2. Inference criteria

The statistical analysis of the pupil data is divided in two levels, a *group-level* analysis and an *individual-level* analysis. On the one hand, the group-level 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. On the other hand, the individual-level analysis 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.

2.4.2.1. Group-level analysis

PDR was modelled as a function of the ordered factor⁴ 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:

- i. rise vs. fall
- ii. 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) test⁵ was implemented to examine this contrast.

2.4.3. Individual-level analysis

For the exploration of the individual cognitive variability effect, we fitted three separate models to examine their interaction with prosodic condition in the three following contrasts:

- i. model 1: neutral vs. rise
- ii. model 2: neutral vs. fall
- iii. model 3: rise vs. fall

⁴ Ordered 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).

⁵ The 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.

In light of the correlation between the cognitive tasks (see Section 2.4.1), and in order to avoid introducing multicollinearity in the models,⁶ only a single measure, the *flanker score* for each participant, was included in models of individual differences. Flanker scores were selected in particular because they likely index inhibitory skills and are therefore highly relevant to any attention-orienting mechanism. In these models, the 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 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 nonlinear 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 nonreference 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 nonreference 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.

3. Results

3.1. Group-level

Figure 6 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 pupil responses to deviants featuring the falling edge tone (falling condition), and the black point-range curve illustrates pupil responses to deviants with neutral intonation (the baseline prosodic condition). Visual inspection of the curves reveals that the different intonational contours modulate the PDR in distinct ways.

⁶ Following a reviewer's suggestion, we conducted additional analyses testing the interaction between prosody and the other two cognitive measures (processing speed and digit span). The results were consistent with the interpretation based on the observed correlations among the cognitive measures. Interested readers can find the full results in the Appendix and the full dataset and analyses in OSF.

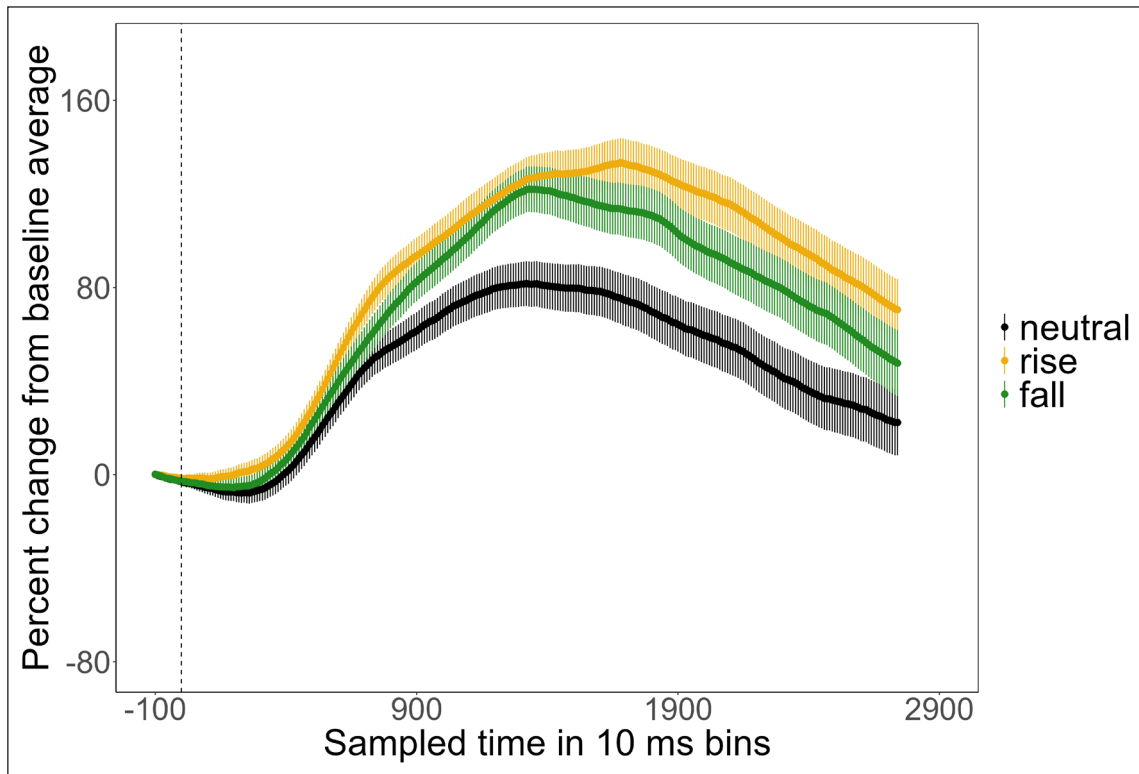


Figure 6: Grand averaged pupil dilation response per prosodic condition over time (sampled in 10 ms bins).

GAMM smooths⁷ (including 95% CIs) of PDR to deviant numbers as a function of prosodic condition are depicted in **Figure 7**. Colour coding of the prosodic conditions' estimated smooths corresponds to the colour coding in **Figure 6**.

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: $\beta = -37.06$, $t = -4.380$, $p < .001$) as well as the *shape* of the PDR curve (smooth difference: $\text{EDF} = 3.198$, $F = 7.038$, $p < .001$), 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 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: $\text{EDF} = 1.048$, $F = 3.632$, $p = .06$), indicating a marginal trend in which rising intonation exhibits subtly a more sustained effect over time.

⁷ GAMM smooths illustrate height and shape properties of the effects, that is, overall mean and latency differences.

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.

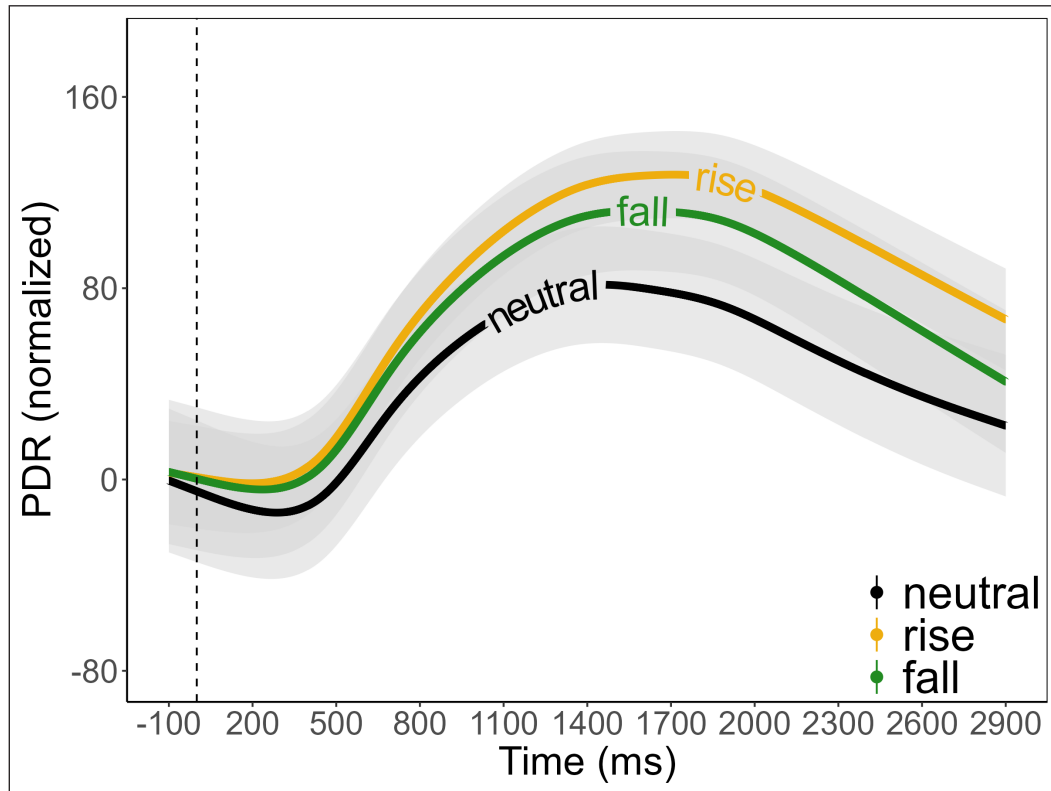


Figure 7: GAMM smooths of pupil dilation response (PDR; normalised) with 95% CIs, across prosodic conditions. PDRs are time-locked to the onset of the deviant (zero ms), as indicated by the vertical dashed line.

3.2. Individual-level

GAMM smooths (including 95% CIs) of PDR to deviant numbers, as a function of prosodic condition and flanker scores, are shown in **Figure 8**. Colour coding of the prosodic conditions' estimated smooths corresponds to the colour coding in previous figures. Flanker scores are illustrated from left (*lower inhibition*) to right (*higher inhibition*) panels. **Figure 9** 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 difference smooth represents the estimated difference between conditions across the range of flanker scores. Negative values on the *x* axis indicate lower inhibition, positive values indicate higher inhibition, and zero represents the average inhibition score. 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.

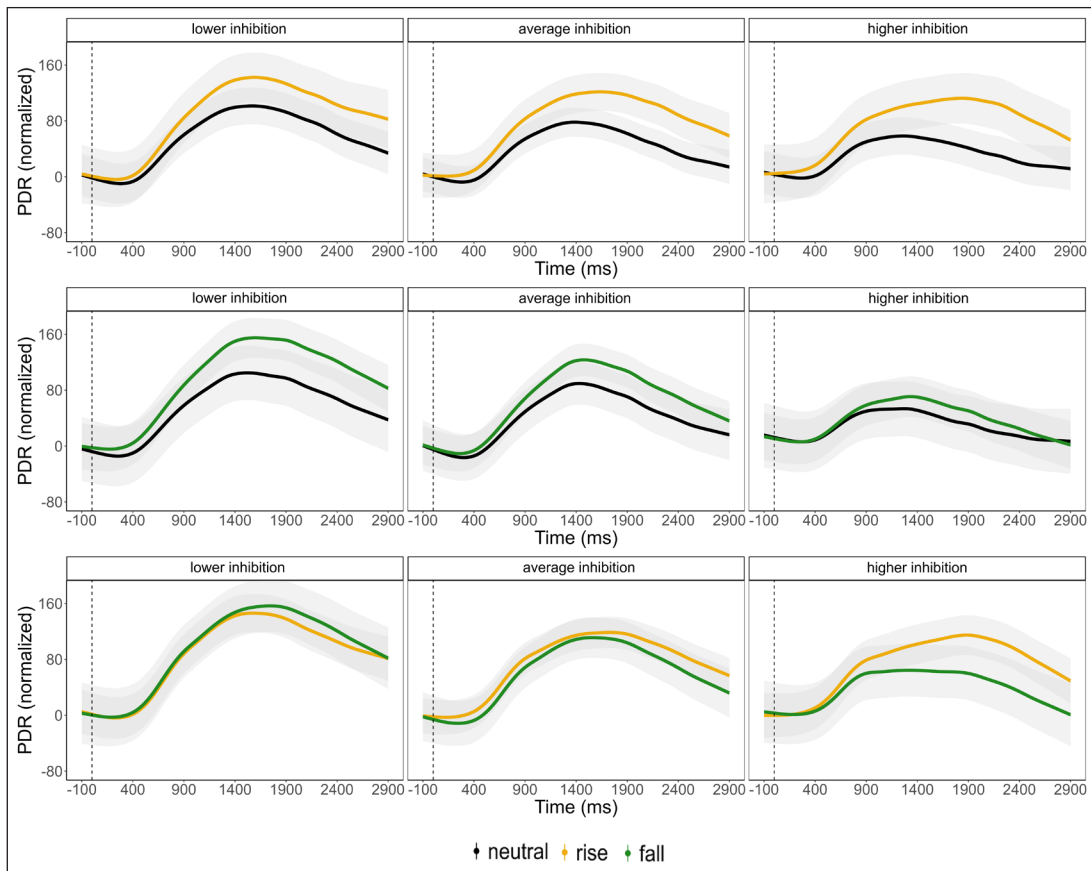


Figure 8: GAMM smooths of pupil dilation response (PDR; normalised) with 95% CIs, across prosodic conditions and flanker scores. Grouping into inhibitory control groups for visual purposes only. PDRs are time-locked to the onset of the deviant (zero ms), as indicated by the vertical dashed line.

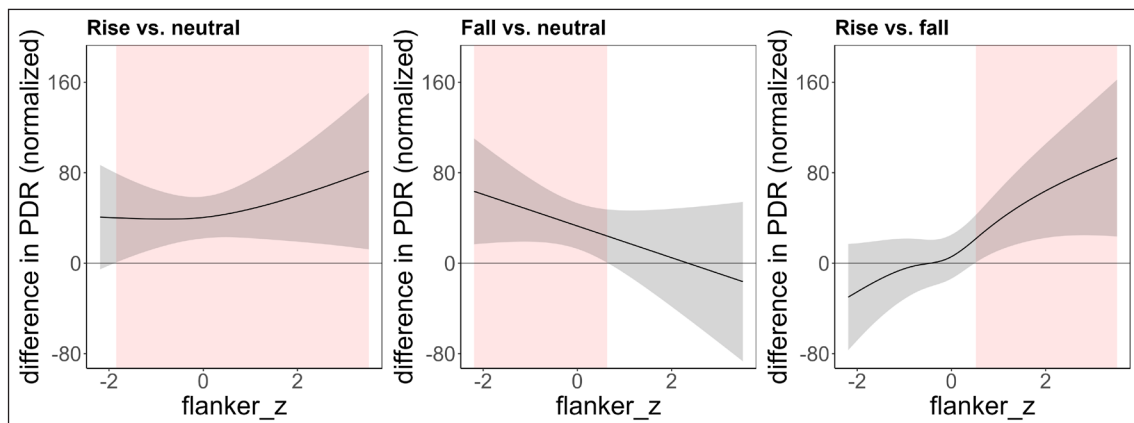


Figure 9: Difference smooths of pupil dilation response PDR (normalised) with 95% CIs for the comparisons among prosodic conditions as a function of flanker scores.

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 8**), 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, $p = .001$). Similarly, for falling intonation (horizontal middle and bottom panels in **Figure 8**), 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, $p = .01$; tensor product smooth: EDF = 10.322, F = 4.581, $p < .001$). For rising intonation (horizontal top and bottom panels in **Figure 8**), 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, $p < .001$).

Comparing neutral to rising intonation (horizontal top panels in **Figure 8** and left panel in **Figure 9**), 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: $\beta = 37.471$, $t = 4.294$, $p < .001$; smooth difference EDF = 5.453, F = 6.164, $p < .001$; tensor product smooth difference: EDF = 6.162, F = 3.346, $p = .001$). Likewise, comparing falling to neutral intonation (horizontal middle panels in **Figure 8** and middle panel in **Figure 9**), 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: $\beta = -23.014$, $t = -2.337$, $p = .01$; smooth difference EDF = 5.170, F = 3.422, $p = .001$; tensor product smooth difference: EDF = 1.003, F = 2.757, $p = .05$). Finally, comparing rising to falling intonation (horizontal bottom panels in **Figure 8** and right panel in **Figure 9**), 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, $p = .05$; tensor product smooth difference: EDF = 7.762, F = 2.887, $p = .001$).

A visible feature of the graphs in **Figure 8** is the late peak in PDR among individuals with higher flanker scores in response to rising deviants. To confirm this effect, we used a permutation test. The difference in mean dilation between the second and first half of the measurements in each trial—after and before 1500 ms respectively—was averaged over the trials per person per condition. The difference in this value between pairs of conditions (rise-neutral, rise-fall, and fall-neutral) was calculated. Participants were divided into two equally-sized groups of 30 each, of higher and lower inhibitory skills. Permutations randomised the relation between participants and the inhibitory skill groups, leaving unchanged the association between participants and the mean PDR difference scores. Of the approximately 10^{17} possible permutations, 10^6 were randomly sampled. The exact p -value calculated varies slightly from run to run, as the permutation simulation process is nondeterministic. The results show a significant effect for the rise-fall difference ($p = .0199$, 99% CI. [0.0195–0.0202]), a near significant p -value for the

rise-neutral difference ($p = .0744$, 99% CI. [0.0738–0.0751]), and nonsignificance for the fall-neutral difference ($p = .745$, 99% CI. [0.744–0.746]).

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]). Furthermore, rising deviants lead to higher PDR later in individuals with better inhibitory skills, compared to falling deviants.

3.3. Summary

To briefly summarise the current findings, at the group level, 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. Within the edge tone conditions, rising intonation resulted in a subtly more sustained response over time than falling intonation.

Considering individual variability, it is evident that inhibitory skill differentially affects PDR modulation across the prosodic conditions. For individuals with strong inhibitory skills, only rising intonation led to more sustained PDRs, while falling and neutral ones evoked transient pupil responses. Furthermore, the effect of the rising deviants occurred later for these participants than for falling and neutral ones, or participants with weak inhibitory skills. 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.

4. Discussion

In this article, we investigated the relevance of domain-final rises and falls (the reflex of High and Low phrase final edge tone respectively) for attention orienting in German. Domain-final rising intonation has been shown to attract attention in serial recall (e.g., Savino et al., 2020; Grice et al., 2024) and ERP experiments (e.g., Lialiou et al., 2024). Here, we 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 twofold. Using PDR as an indicator of attention orienting, the main objective was to test whether deviant numbers in highly predictable numerical sequences capture more attention, and thereby evoking more robust PDRs, when realised with a final rise, as compared to when such sequences are produced with a final fall or when their intonation does not differ from that of the standards (the neutral case). In addition, we explored whether variation

in cognitive measures subserving attention orienting would also associate with differences in responses to numerical deviants.

4.1. Group-level effects and their implications

Starting with the group-level findings of this study, the present results provide evidence in favour of our first hypothesis: deviants in a numerical sequence do indeed elicit a PDR, regardless of intonation. These results corroborate previous findings on auditory cognition claiming that attention orienting is substantiated 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, produced with 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 (i.e., changing) acoustic properties attract attention when presented as deviants (e.g., Rinne et al., 2005, 2006; Bach et al., 2008; Macdonald & Campbell, 2011).

Moving to our second hypothesis, and hence the comparison between rising and falling edge tones, a subtle difference in PDR *shape* was found between the two tones, which suggests that rising tones evoked subtly more long-lasting PDRs than falling tones. We interpret the difference in PDR shape to indicate that a more sustained attentional response has been allocated to deviants marked by the former than to those marked by the latter tones. Nonetheless, the individual-level results suggested that this difference depends on the cognitive bandwidth of the individual. In what follows, we first discuss the contribution of individual cognitive variability in attention orienting. We then discuss the function of edge tones as attention-orienting devices across different cognitive bandwidths.

4.2. Individual variability contributes to attention orienting

As mentioned, cognitive variability was measured on the basis of three cognitive skills: processing speed, inhibitory ability, and working memory (WM) resistance. As the measurements were weakly correlated with one another (see Section 2.4.1), statistical analysis of PDR included only flanker scores as an interactive predictor, which was reasoned to best index inhibitory ability.

The data show that the presentation of deviants attracts listeners' involuntary attentional resources, regardless of intonation. This is manifested in the dilations of 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 to more rapid PDRs to deviants produced with falling or baseline neutral intonation. Conversely, individuals with weaker inhibitory ability were less sensitive to differences in edge tones; they responded to both rising and falling deviant prosodic realisations with equally sustained PDRs, generating an increased and prolonged effect compared to the neutral baseline.

As mentioned in Section 1.1, according to Strauch et al. (2022), the time course of pupil response is indicative of the different stages of the orienting response; transient dilations indicate preattentive processes, whereas more prolonged dilations reflect processes available to the conscious level. Under this interpretation of pupil response, the present data are compatible with a multistage or multi-process account. The presentation of a deviant initially results in an involuntary attention switch across all prosodic conditions. After the switch, however, other top-down executive operations such as processing speed, inhibitory ability, and WM, cooperate and continuously interact with the attentional system. The attentional system, in turn, determines whether the preattentive involuntary attention switch will lead to full awareness of the deviant, or whether the auditory processing system will continue with the processing of the following input.

Inhibitory ability is a crucial cognitive factor to consider in understanding the observed PDR variability across individuals. Nonetheless, it is worth considering all three cognitive skills comprising cognitive profiles—inhibitory ability, processing speed, and WM resistance—together in an attempt 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 a higher WM resistance (resulting from a smaller WM capacity). Conversely, individuals with poorer inhibitory abilities tended to show a faster processing speed and a lower WM resistance (resulting from a larger WM capacity). For ease, hereafter we call the first *strong inhibitors*, and the latter, *weak inhibitors*. Let us now consider the aforementioned PDR findings in light of individual cognitive variability.

Strong inhibitors 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 a longer time, indicating that more attentional resources have been allocated to those deviants. Another possible explanation could be that the perception of the relevant deviant has been brought to a more conscious level, yet this remains to be further investigated. 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 acoustic signal. In the case of a rising deviant, the processing system affords it high importance, due to the deviant's high prominence. Given the established importance of the deviant, inhibition is blocked. The deviant thus enters WM 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 low prosodic 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 of attentional control have shown that increased WM load prevents (or at least attenuates) auditory distraction by utilising a top-down control (e.g., SanMiguel et al., 2008). Here, individuals' high WM resistance (resulting from a low WM capacity) potentially leads to high WM load 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 strong inhibitors, weak inhibitors 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), managed to break through the shield of voluntary attention. However, when a deviant did not differ prosodically from the standard numbers, owing to its very low prosodic prominence, it evoked only an involuntary switch, drawing attention back to the initial focus. As the data reported in this article show, these individuals exhibited lower inhibitory skills and further tended to be characterised by faster processing speed and lower WM resistance (related larger WM capacity). 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 based on its prosodic prominence. It appears that for this group of individuals, due to the fast processing and the inefficient evaluation of the prominence value of the deviant, in conjunction with a weaker inhibitory mechanism and low WM resistance, the strength of the boundary that marks the deviant events suffices to enable them to move to later stages of orienting.

As mentioned in Section 1.3, research on individual variability in auditory attention is sparse, and, to our knowledge, no previous study has directly tested the effect of inhibition or processing speed on auditory attention. With regard to WM, the present results are compatible with Berti & Schröger (2003) and SanMiguel et al. (2008) in reporting that increased WM load, and thus high WM resistance, attenuates or even prevents auditory distraction. Nonetheless, the present findings stand in contrast with other studies reporting that listeners with high WM capacity 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 increased WM load (corresponding to high WM resistance) 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 WM capacity 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, we take these findings as further evidence in support of the present results. Finally, with regard to processing speed and WM resistance, our results are in line with a study on visual attention by Heitz & Engle (2007) reporting that individuals with large WM capacity (and thus small WM resistance) had faster responses than individuals with small WM capacity.

4.3. The transition from preattentive to later stages of processing is visible in the PDR

Let us now focus on the comparison between rising and falling edge tones (i.e., our second research question). Our results indicated different patterns as a function of cognitive profile. On the one hand, for strong inhibitors (i.e., individuals with strong suppressive mechanisms accompanied by slow processing speed), a PDR *shape* difference was found between the two tones. This difference indicates that rising tones evoked more long-lasting PDRs than falling tones. On the other hand, for weak inhibitors (i.e., individuals with weak suppressive mechanisms accompanied by fast processing speed), we found no PDR difference between the two tones, in that they both exhibited most sustained responses in comparison to strong inhibitors. These results potentially reflect that individuals with strong inhibitory capacity, as opposed to those with weaker inhibitory capacity, were better able to use pitch-related information from the signal, which helped guide their allocation of attention. This, in turn, partially determined the interaction with additional top-down mechanisms used to recover from the deviance.

It has been suggested that the latency of attention-related PDRs is indicative of the processing stage in which the attentional mechanism is activated: transient dilations reflect preattentive attentional processes, while more prolonged responses indicate processes available for conscious processing (see Strauch et al., 2022). Whilst attention-related PDRs are usually transitory and evoked by sensory events such as an auditory deviance, top-down processes can also affect PDRs, in some cases prolonging pupillary response (Strauch et al., 2022). Thus, an attention-related PDR, elicited 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, possibly bringing the deviant into awareness. In other words, the route from preattentive to later processing stages of a deviant is reflected in the time course of pupil responses.

For strong inhibitors, 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 predictable pattern produced with neutral intonation. Nonetheless, when the deviant number is produced with rising pitch, this elicited prolonged pupil dilations, indicating that, in this case, the violation first causes an involuntary attention switch, followed by voluntary attention orienting. 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 later processing stages, in contrast to falling (and neutral) deviants, jolting participants out of their inhibition, can be attributed to their prominence. These results are 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. Further, 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).

For weak inhibitors, the observation is that the presentation of the deviant, featuring either of the two edge tones, evokes not only an increased but also a prolonged PDR compared to a neutral deviant. This means that participants with weak suppressive mechanisms consume more attentional resources towards the word bearing both rising and falling edge tones marking a deviant after the initial involuntary switch. This could be potentially due to the same function of rising and falling edge tones in this study, namely marking the end of a smaller or a larger unit. The importance of the context for attentional resources has been manifested in Lialiou et al. (2024), where the authors showed that involuntary attention orienting is signal-driven, while voluntary attention is driven by meaningful aspects of intonation licensed by contextually created expectations.

Taken together, these findings suggest that voluntary attention is guided by different prosodic cues depending on individuals' inhibitory control. For strong inhibitors, the direction of pitch movement (rising vs. falling) appears to drive voluntary attention, whereas for weak inhibitors it is the strength of the prosodic boundary (strong vs. weak) that serves as the guiding cue. This distinction highlights how individual differences in inhibition shape the use of prosodic information in attention orienting.

4.4. Implications for intonational theory

We argue that the present results highlight the attention-orienting function of edge tones, suggesting that domain-final rises and falls on deviant stimuli also enhance the ability of the deviant to attract attention. These findings thus strengthen the case for the role of intonation at the edges of constituents in attention orienting. As discussed in Section 1.2, 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 attributed a mere phrasing function. In that sense, it has been claimed that pitch accents are better in directing listener attention than edge tones. This has already been called into question by results from two serial recall studies (e.g., Savino et al., 2020; Grice et al., 2024) which report that rising edge tones

marking the final item of nonfinal triplets boost the recall accuracy of the whole triplet, hence orienting attention to the whole domain. The work reported in Lialiou et al. (2024) adds to these 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 present study contribute further to this discovery that tonal events associated with the edges of constituents can also direct listener attention. Building on the aforementioned serial recall results, one speculation is that in individuals with strong suppressive mechanisms, 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. For individuals with weak suppressive mechanisms, it appears that the strength of the boundary marking the deviant (i.e., the digit the listener is concerned with at the moment) is enough for attracting attention. In terms of information packaging, a strong boundary (like our rising and falling edge tone on the deviants) means the end of list, wrapping up, whereas a weaker boundary (like the one in neutral deviants) means more to come. It is hence possible that due to the fast processing and the weak suppressive mechanisms that characterises those individuals, the specific tone (and its prosodic prominence value) is additional information that the processing system cannot handle or does not need. However, the strength of the boundary, flagging a prominent position like the end of a subunit, encourages the listener to integrate the deviant with previous items in the same domain, increasing listener awareness towards the violation.

In a discourse scenario, individuals with weak suppressive mechanisms may be less effective at inhibiting alternative activations while predicting turn endings, whereas individuals with stronger suppressive mechanisms might generate a narrower set of predictions. If a strong boundary (especially a rising one, for some individuals) prompts the listener to integrate information within a constituent to a greater degree than a weak boundary would, we could speculate that the strong boundary boosts cognitive activation within that constituent, making a referent given, such that it can be referred back to via anaphoric devices or deaccentuation. This potential implication for discourse processing would need to be empirically tested.

5. Conclusion

The present study 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), we investigated whether domain-final rising pitch in speech takes on a special role in attention orienting by measuring listeners' pupil dilation response (PDR). Further, we focused on the contribution of individual cognitive variability to attention orienting.

The present findings can be summarised as follows:

- i. Attention orienting is corroborated by an expectancy violation mechanism.
- ii. Pupillometry is a rigorous technique for studying attention orienting.
- iii. Intonation takes on a special role in attracting attention, even when it does not involve prominence cueing in traditional terms.
- iv. 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.

Considering the present findings holistically, this study is highly compatible with the notion that individuals differ in the cognitive mechanisms they have at their disposal. The evaluation of deviants by the processing system constitutes the first crucial step in determining which operations are activated during the attention-orienting and resolution process. Individuals may or may not have an efficient processing system, leading to successful or unsuccessful evaluation, respectively. The next step involves the activation of suppressive mechanisms such as inhibition and WM resistance. Likewise, individuals may or may not have efficient suppressive mechanisms to protect their attentional system from potential overloading. In this study, some individuals, who we have called strong inhibitors, appear to have at their disposal both an efficient processing system, which evaluates deviants' importance more successfully, and sufficient suppressive mechanisms, activated when needed. For the remaining (weak inhibitor) individuals, we have hypothesized that they have neither enough time at their disposal to evaluate the prosodic prominence value of the deviant nor strong enough suppression mechanisms to do so. As a result, deviants featuring a strong boundary tone, regardless of whether this boundary tone is a rise or a fall, are sufficient to command additional attentional resources, which potentially maximises attentional processing load.

To conclude, the present results highlight the attention-orienting function of edge tones, suggesting that domain-final rises (and falls for some individuals) on deviant stimuli also enhance the ability of the deviant to attract attention. That edge tones 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 in a language like German. The present experimental investigation corroborates the findings of recent studies which point towards an attention-orienting function of edge tones (e.g., Savino et al., 2020; Grice et al., 2024; Lialiou et al., 2024), by showing that not only accentual but also edge tone-related intonational contours 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 role of edge tones in the prominence-cueing function.

Appendix

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
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

(Contd.)

Participant ID	Flanker	Odd-man-out	Digit Span
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
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

(Contd.)

Participant ID	Flanker	Odd-man-out	Digit Span
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

Table A1: Mean performance per participant across the three cognitive tasks.

A.1 Results – Processing speed

The models indicated an interaction between prosodic condition and odd-man-out scores. More specifically, the results show that PDR changes as a function of processing speed across prosodic conditions. For neutral intonation (horizontal top [and middle] panels in **Figure A1**), it was shown that processing speed modulated PDR *shape*, such that the slower the speed, the longer the PDR (tensor product smooth: EDF = 5.997, $F = 2.171$, $p = .01$). Similarly, for falling intonation (horizontal middle [and bottom] panels in **Figure A1**), processing speed also affected PDR *shape*. From fast to average to slow processing speed, a decrease in the duration of the PDR was observed (tensor product smooth: EDF = 7.748, $F = 8.597$, $p < .0001$). For rising intonation (horizontal [top and] bottom panels in **Figure A1**), the results also show that PDR *shape* changed as a function of processing speed, such that the slower the speed, the longer the PDR (tensor product smooth: EDF = 8.179, $F = 3.055$, $p = .001$).

Comparing neutral to rising intonation (horizontal top panels in **Figure A1**), PDRs differed in both *height* and *shape*, as a function of processing speed: from faster to slower processing speed, the slower the speed, the stronger the rising PDRs, showing an increased and long-lasting effect (parametric difference: $\beta = 37.585$, $t = 4.369$, $p < .0001$; tensor product smooth difference: $\text{EDF} = 8.179$, $F = 3.055$, $p = .001$). Likewise, comparing falling to neutral intonation (horizontal middle panels in **Figure A1**), PDRs differed in both *height* and *shape* as a function of fast processing speed, such that the faster the processing speed, the weaker the neutral PDRs, showing a decreased and subtly faster effect (parametric difference: $\beta = -23.798$, $t = -2.381$, $p = .01$; tensor product smooth difference: $\text{EDF} = 5.992$, $F = 1.915$, $p = .05$). Finally, comparing rising to falling intonation (horizontal bottom panels in **Figure A1**), PDRs differed only in *shape* as a function of slow processing speed, such that the slower the speed, the faster the falling PDRs (tensor product smooth difference: $\text{EDF} = 4.640$, $F = 1.915$, $p = .05$).

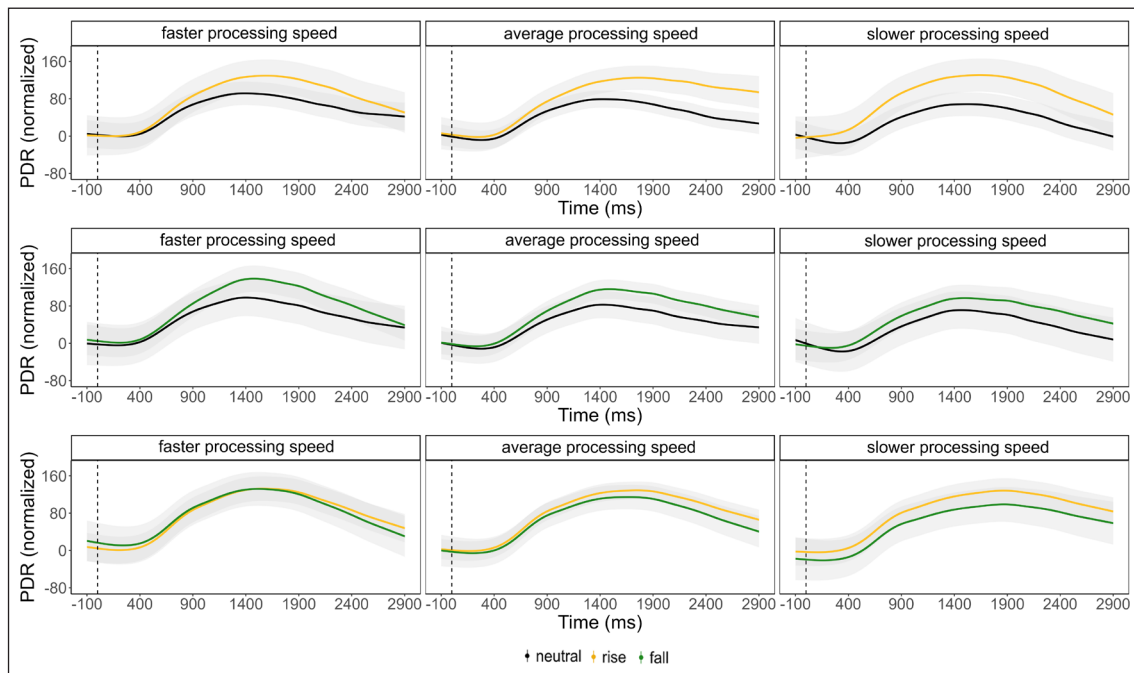


Figure A1: GAMM smooths of PDR (normalised) with 95% CIs, across prosodic conditions and odd-man-out scores. Grouping into processing speed groups for visual purposes only. PDRs are time-locked to the onset of the deviant (zero ms), as indicated by the vertical dashed line.

A.2 Results – Digit Span

The models indicated an interaction between prosodic condition and digit span scores. More specifically, the results show that PDR changes as a function of digit span across prosodic conditions. For neutral intonation (horizontal top [and middle] panels in **Figure A2**), it was shown that digit span modulated PDR *shape*, such that the larger the span, the longer the PDR (tensor product smooth: $\text{EDF} = 11.075$, $F = 5.279$, $p < .0001$). Similarly, for falling intonation

(horizontal middle [and bottom] panels in **Figure A2**), digit span also affected PDR *shape*. From smaller to average to larger digit span, a decrease in the duration of the PDR was observed (difference smooth: EDF = 4.338, $F = 5.895$, $p < .0001$; tensor product smooth: EDF = 12.415, $F = 10.139$, $p < .0001$). For rising intonation (horizontal [top and] bottom panels in **Figure A2**), the results also show that PDR *shape* changed as a function of digit span, such that the smaller the span score, the longer the PDR (tensor product smooth: EDF = 11.945, $F = 5.193$, $p < .0001$).

Comparing neutral to rising intonation (horizontal top panels in **Figure A2**), PDRs differed in both *height* and *shape*, as a function of digit span: from smaller to larger digit span, the smaller the span, the stronger the rising PDRs, showing an increased and long-lasting effect (parametric difference: $\beta = 35.022$, $t = 4.027$, $p < .0001$; smooth difference: EDF = 2.914, $F = 5.624$, $p < .001$; tensor product smooth difference: EDF = 11.945, $F = 5.193$, $p < .0001$). Likewise, comparing falling to neutral intonation (horizontal middle panels in **Figure A2**), PDRs differed in both *height* and *shape* as a function of low digit span, such that the lower the digit span, the weaker the neutral PDRs, showing a decreased and subtly faster effect (parametric difference: $\beta = -46.02$, $t = -3.678$, $p = .001$; tensor product smooth difference: EDF = 13.001, $F = 11.160$, $p < .0001$). Finally, comparing rising to falling intonation (horizontal bottom panels in **Figure A2**), PDRs differed only in *shape* as a function of large digit span, such that the larger the span score, the faster the falling PDRs (tensor product smooth difference: EDF = 3.738, $F = 13.483$, $p < .0001$).

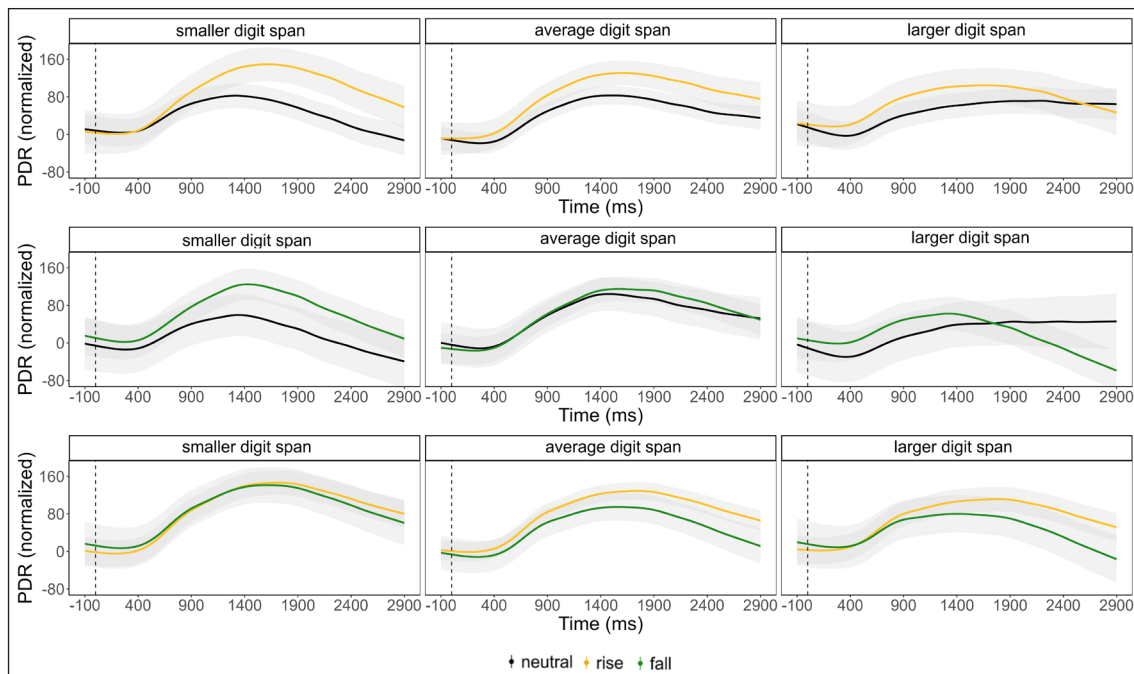


Figure A2: GAMM smooths of PDR (normalised) with 95% CIs, across prosodic conditions and digit span scores. Grouping into span groups for visual purposes only. PDRs are time-locked to the onset of the deviant (zero ms), as indicated by the vertical dashed line.

Data accessibility statement

Data and scripts for all analyses have been made available online at OSF platform (<https://osf.io/j8295/overview>).

Acknowledgements

We would like to thank Brita Rietdorf and Claudia Kilter for their excellent support in recruiting participants and running the experiment. We also thank Solveigh Janzen for her help with annotating the speech data and Christine Röhr for recording our speech material. We are also very grateful to Márton Sóskuthy for his invaluable feedback on our GAMM analyses. We furthermore thank Heiko Seeliger, the participants of the IfL-Phonetik colloquium and the TAI 2025 conference for their discussion and feedback. Last but not least, we thank all of our participants—without them, this study would not have been possible.

Funding information

The research for this article has been funded by the Deutsche Forschungsgemeinschaft (German Research Foundation; <https://doi.org/10.13039/501100001659>), grant number: Project-ID 281511265 – SFB 1252 “Prominence in Language” in the project A01 “Intonation and attention orienting: Neurophysiological and behavioural correlates” at the University of Cologne.

Competing interests

The authors have no competing interests to declare.

Author contributions

Conceptualisation: ML, JH, PBS, MG; Methodology: ML, JH, TME, PBS, MG; Software: ML, TME; Formal analysis: ML, TME; Investigation: ML; Resources: PBS, MG; Data curation: ML; Writing – original draft: ML; Writing – review and editing: ML, JH, TME, PBS, MG; Visualisation: ML; Supervision: JH, PBS, MG; Project administration: ML, PBS, MG; Funding acquisition: PBS, MG.

Author Note

Parts of this article are based on material from the first author’s doctoral dissertation.

References

- Alain, C., Woods, D. L., & Ogawa, K. H. (1994). Brain indices of automatic pattern processing. *Neuroreport*, 6(1), 140–144. <https://doi.org/10.1097/00001756-199412300-00036>
- Alamia, A., VanRullen, R., Pasqualotto, E., Mouraux, A., & Zenon, A. (2019). Pupil-linked arousal responds to unconscious surprisal. *The Journal of Neuroscience*, 39(27), 5369–5376. <https://doi.org/10.1523/JNEUROSCI.3010-18.2019>

- Albert, A. (2023). *A model of sonority based on pitch intelligibility*. Language Science Press. <https://doi.org/10.5281/zenodo.7837176>
- Albert, A., Cangemi, F., & Grice, M. (2018). Using periodic energy to enrich acoustic representations of pitch in speech: A demonstration. *Proceedings of Speech Prosody 2018*, 804–808. <https://doi.org/10.21437/SpeechProsody.2018-162>
- Albert, A., Cangemi, F., Grice, M., & Ellison, T. M. (2020). *ProPer: PROsodic analysis with PERiodic energy* [Computer software]. OSF. <https://osf.io/28ea5/>
- Arvaniti, A., Grice, M., & D’Imperio, M. (2025). Advancements of phonetics in the 21st century: Intonation. *Journal of Phonetics*, 113, 101459. <https://doi.org/10.1016/j.wocn.2025.101459>
- Arvaniti, A., & Ladd, D. R. (2023). Prosodic prominence across languages. *Annual Review of Linguistics*, 9, 171–193. <https://doi.org/10.1146/annurev-linguistics-031120-101954>
- Bach, D. R., Schachinger, H., Neuhoff, J. G., Esposito, F., Salle, F. D., Lehmann, C., Herdener, M., Scheffler, K., & Seifritz, E. (2008). Rising Sound Intensity: An Intrinsic Warning Cue Activating the Amygdala. *Cerebral Cortex*, 18(1), 145–150. <https://doi.org/10.1093/cercor/bhm040>
- Baumann, S., & Röhr, C. T. (2015). The perceptual prominence of pitch accent types in German. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*, 1–5.
- Baumann, S., & Trouvain, J. (2001). On the prosody of German telephone numbers. *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech 2001)*, 557–560. <https://doi.org/10.21437/Eurospeech.2001-149>
- Baumann, S., & Winter, B. (2018). What makes a word prominent? Predicting untrained German listeners’ perceptual judgments. *Journal of Phonetics*, 70, 20–38. <https://doi.org/10.1016/j.wocn.2018.05.004>
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91(2), 276–292. <https://doi.org/10.1037/0033-2909.91.2.276>
- Beckman, M. E. (2012). *Stress and Non-Stress Accent*. De Gruyter Mouton. <https://doi.org/10.1515/9783110874020>
- Berti, S., & Schröger, E. (2003). Working memory controls involuntary attention switching: Evidence from an auditory distraction paradigm. *The European Journal of Neuroscience*, 17(5), 1119–1122. <https://doi.org/10.1046/j.1460-9568.2003.02527.x>
- Bishop, J. (2012). WPP, no.111: Focus, prosody, and individual differences in “autistic” traits: Evidence from cross-modal semantic priming. *UCLA Work. Papers in Phonetics*, 111, 1–26.
- Bishop, J. (2016). Individual differences in top-down and bottom-up prominence perception. *Proc. Speech Prosody 2016*, 668–672. <https://doi:10.21437/SpeechProsody.2016-137>
- Bishop, J., Kuo, G., & Kim, B. (2020). Phonology, phonetics, and signal-extrinsic factors in the perception of prosodic prominence: Evidence from Rapid Prosody Transcription. *Journal of Phonetics*, 82, 100977. <https://doi.org/10.1016/j.wocn.2020.100977>
- Boersma, P., & Weenink, D. (2024). *Praat: Doing phonetics by computer* (Version 6.4.06) [Computer software]. <http://www.praat.org/>
- Cangemi, F. (2015). *Mausmooth* [Computer software]. <https://ifl.phil-fak.uni-koeln.de/sites/linguistik/Phonetik/pdf-publications/2015/cangemi2015mausmooth.pdf>

- Cangemi, F., Albert, A., & Grice, M. (2019). Modelling intonation: Beyond segments and tonal targets. *Proceedings of the 19th International Congress of Phonetic Sciences*, 572–576.
- Cangemi, F., & Baumann, S. (2020). Integrating phonetics and phonology in the study of linguistic prominence. *Journal of Phonetics*, 81, 100993. <https://doi.org/10.1016/j.wocn.2020.100993>
- Chafe, W. L. (1974). Language and Consciousness. *Language*, 50(1), 111. <https://doi.org/10.2307/412014>
- Chen, A. (2003). Language Dependence in Continuation Intonation. *Proceedings of the 15th International Congress of Phonetic Sciences*, 1069–1072.
- Chobert, J., François, C., Habib, M., & Besson, M. (2012). Deficit in the preattentive processing of syllabic duration and VOT in children with dyslexia. *Neuropsychologia*, 50(8), 2044–2055. <https://doi.org/10.1016/j.neuropsychologia.2012.05.004>
- Cole, J., Hualde, J. I., Smith, C. L., Eager, C., Mahrt, T., & de Souza, R. N. (2019). Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish. *Journal of Phonetics*, 75, 113–147. <https://doi.org/10.1016/j.wocn.2019.05.002>
- Cowan, N. (1998). *Attention and Memory: An Integrated Framework*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195119107.001.0001>
- Dimberg, U. (1990). Facial electromyography and emotional reactions. *Psychophysiology*, 27(5), 481–494. <https://doi.org/10.1111/j.1469-8986.1990.tb01962.x>
- Dingemans, M., Torreira, F., & Enfield, N. J. (2013). Is “Huh?” a Universal Word? Conversational Infrastructure and the Convergent Evolution of Linguistic Items. *PLoS ONE*, 8(11), e78273. <https://doi.org/10.1371/journal.pone.0078273>
- Doeller, C. F., Opitz, B., Mecklinger, A., Krick, C., Reith, W., & Schröger, E. (2003). Prefrontal cortex involvement in preattentive auditory deviance detection: Neuroimaging and electrophysiological evidence. *NeuroImage*, 20(2), 1270–1282. [https://doi.org/10.1016/S1053-8119\(03\)00389-6](https://doi.org/10.1016/S1053-8119(03)00389-6)
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, 16(1), 143–149. <https://doi.org/10.3758/BF03203267>
- Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural Mechanisms of Involuntary Attention to Acoustic Novelty and Change. *Journal of Cognitive Neuroscience*, 10(5), 590–604. <https://doi.org/10.1162/089892998562997>
- Frearson, W., & Eysenck, H. J. (1986). Intelligence, reaction time (RT) and a new “odd-man-out” RT paradigm. *Personality and Individual Differences*, 7(6), 807–817. [https://doi.org/10.1016/0191-8869\(86\)90079-6](https://doi.org/10.1016/0191-8869(86)90079-6)
- Frischkorn, G. T., Wilhelm, O., & Oberauer, K. (2022). Process-oriented intelligence research: A review from the cognitive perspective. *Intelligence*, 94, 101681. <https://doi.org/10.1016/j.intell.2022.101681>
- Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138. <https://doi.org/10.1038/nrn2787>
- Friston, K. (2018). Does predictive coding have a future? *Nature Neuroscience*, 21(8), 1019–1021. <https://doi.org/10.1038/s41593-018-0200-7>

- Grabe, E. (1998). *Comparative Intonational Phonology: English and German* [Doctoral dissertation, Radboud University Nijmegen]. MPG PuRe. https://pure.mpg.de/view/item_2057683
- Graham, F. K., & Clifton, R. K. (1966). Heart-rate change as a component of the orienting response. *Psychological Bulletin*, 65(5), 305–320. <https://doi.org/10.1037/h0023258>
- Grice, M. (2022). Autosegmental-metrical phonology – Unpacking the boxes. *Zeitschrift Für Sprachwissenschaft*, 41(2), 393–411. <https://doi.org/10.1515/zfs-2022-2002>
- Grice, M., Baumann, S., & Benz Müller, R. (2005). German Intonation in Autosegmental-Metrical Phonology. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 55–83). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199249633.003.0003>
- Grice, M., & Kügler, F. (2021). Prosodic Prominence – A Cross-Linguistic Perspective. *Language and Speech*, 64(2), 253–260. <https://doi.org/10.1177/00238309211015768>
- Grice, M., Ritter, S., Niemann, H., & Roettger, T. B. (2017). Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics*, 64, 90–107. <https://doi.org/10.1016/j.wocn.2017.03.003>
- Grice, M., Savino, M., Schumacher, P. B., Röhr, C. T., & Ellison, T. M. (2024). *Rises on Pitch Accents and Edge Tones Affect Serial Recall Performance at Item and Domain levels*. OSF. <https://doi.org/10.31234/osf.io/tq6jp>
- Grice, M., Wehrle, S., Krüger, M., Spaniol, M., Cangemi, F., & Vogeley, K. (2023). Linguistic prosody in autism spectrum disorder—An overview. *Language and Linguistics Compass*, 17(5), e12498. <https://doi.org/10.1111/lnc3.12498>
- Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511616983>
- Gussenhoven, C., Repp, B. H., Rietveld, T., Rump, H. H., & Terken, J. (1997). The perceptual prominence of fundamental frequency peaks. *The Journal of the Acoustical Society of America*, 102(5), 3009–3022.
- Gussenhoven, C., & Rietveld, T. (1988). Fundamental frequency declination in Dutch: Testing three hypotheses. *16(3)*, 335–369. [https://doi.org/10.1016/S0095-4470\(19\)30509-1](https://doi.org/10.1016/S0095-4470(19)30509-1)
- Hart, J. T., Collier, R., & Cohen, A. (1990). *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody*. Cambridge University Press.
- Heitz, R. P., & Engle, R. W. (2007). Focusing the spotlight: Individual differences in visual attention control. *Journal of Experimental Psychology. General*, 136(2), 217–240. <https://doi.org/10.1037/0096-3445.136.2.217>
- Hsu, C.-H., Evans, J. P., & Lee, C.-Y. (2015). Brain responses to spoken F0 changes: Is H special? *Journal of Phonetics*, 51, 82–92. <https://doi.org/10.1016/j.wocn.2015.02.003>
- Hughes, R. W. (2014). Auditory distraction: A duplex-mechanism account. *PsyCh Journal*, 3(1), 30–41. <https://doi.org/10.1002/pchj.44>
- Hughes, R. W., & Jones, D. M. (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. <https://doi.org/10.1037/0096-1523.29.1.199>

- Hughes, R. W., & Jones, D. M. (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. <https://doi.org/10.1037/0096-1523.31.2.316>
- Hughes, R. W., Vachon, F., & Jones, D. M. (2007). Disruption of short-term memory by changing and deviant sounds: Support for a duplex-mechanism account of auditory distraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(6), 1050–1061. <https://doi.org/10.1037/0278-7393.33.6.1050>
- Hurley, R., & Bishop, J. (2016). Prosodic and individual influences on the interpretation of “only.” *Speech Prosody 2016*, 193–197. <https://doi.org/10.21437/SpeechProsody.2016-40>
- Johansson, B., & Balkenius, C. (2018). A computational model of pupil dilation. *Connection Science*, 30(1), 5–19. <https://doi.org/10.1080/09540091.2016.1271401>
- Joshi, S., & Gold, J. I. (2020). Pupil Size as a Window on Neural Substrates of Cognition. *Trends in Cognitive Sciences*, 24(6), 466–480. <https://doi.org/10.1016/j.tics.2020.03.005>
- Jun, S.-A. (2014). Prosodic Typology: By Prominence Type, Word prosody, and Macro-rhythm. In S.-A. Jun (Ed.), *Prosodic Typology II: The Phonology of Intonation and Phrasing* (pp. 520–539). Oxford University Press.
- Jun, S.-A., & Bishop, J. (2015a). Priming Implicit Prosody: Prosodic Boundaries and Individual Differences. *Language and Speech*, 58(4), 459–473. <https://doi.org/10.1177/0023830914563368>
- Jun, S.-A., & Bishop, J. (2015b). Prominence in relative clause attachment: Evidence from prosodic priming. In L. Frazier & E. Gibson (Eds.), *Explicit and implicit prosody in sentence processing: Studies in honor of Janet Dean Fodor* (pp. 217–240). Cham: Springer International Publishing.
- Kember, H., Choi, J., Yu, J., & Cutler, A. (2021). The Processing of Linguistic Prominence. *Language and Speech*, 64(2), 413–436. <https://doi.org/10.1177/0023830919880217>
- Keye, D., Wilhelm, O., Oberauer, K., & van Ravenzwaaij, D. (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. <https://doi.org/10.1007/s00426-008-0188-9>
- Knight, R.-A. (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. <https://doi.org/10.1177/0023830908098541>
- Kohler, K., & Gartenberg, R. (1991). The perception of accents: F0 peak height versus F0 peak position. *Proceedings of AIPUK*, 25, 219–242.
- Kyröläinen, A.-J., Porretta, V., Rij, J. van, & Järviö, J. (2020). *PupilPre: Preprocessing Pupil Size Data* (Version 0.6.2) [Computer software]. <https://cran.r-project.org/web/packages/PupilPre/index.html>
- Ladd, D. R. (2008). *Intonational Phonology* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511808814>
- Ladd, D. R., & Morton, R. (1997). The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics*, 25(3), 313–342. <https://doi.org/10.1006/jpho.1997.0046>

- Leiner, D. J. (2024). *SoSci Survey* (Version 3.5.02i) [Computer software]. www.sosicisurvey.de
- Lialiou, M., Grice, M., Röhr, C. T., & Schumacher, P. B. (2024). Auditory Processing of Intonational Rises and Falls in German: Rises Are Special in Attention Orienting. *Journal of Cognitive Neuroscience*, 1–24. https://doi.org/10.1162/jocn_a_02129
- Lialiou, M., Grice, M., & Schumacher, P. B. (2025). A test battery for measuring individual cognitive variability. [Computer software]. <https://doi.org/10.17605/OSF.IO/MUH9T>
- Lialiou, M., Grice, M., & Schumacher, P. B. (in press). A test battery for measuring individual cognitive ability: A brief practical tutorial [Author Accepted Manuscript]. *Europe's Journal of Psychology*. <https://doi.org/10.23668/psycharchives.21626>
- Liao, H.-I., Yoneya, M., Kidani, S., Kashino, M., & Furukawa, S. (2016). Human Pupillary Dilation Response to Deviant Auditory Stimuli: Effects of Stimulus Properties and Voluntary Attention. *Frontiers in Neuroscience*, 10. <https://doi.org/10.3389/fnins.2016.00043>
- Lieberman, M. Y. (1975). *The intonational system of English* [PhD Thesis, Massachusetts Institute of Technology]. <https://dspace.mit.edu/handle/1721.1/27376>
- Lorenzen, J., Roessig, S., & Baumann, S. (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. <https://doi.org/10.3389/fpsyg.2024.1296933>
- Macdonald, M., & Campbell, K. (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. <https://doi.org/10.1016/j.bandc.2011.08.014>
- Maltzman, I. (1979). Orienting Reflexes and Classical Conditioning in Humans. In *The Orienting Reflex in Humans*. Routledge.
- Marois, A., Labonté, K., Parent, M., & Vachon, F. (2018). Eyes have ears: Indexing the orienting response to sound using pupillometry. *International Journal of Psychophysiology*, 123, 152–162. <https://doi.org/10.1016/j.ijpsycho.2017.09.016>
- Marois, A., Marsh, J. E., & Vachon, F. (2019). Is auditory distraction by changing-state and deviant sounds underpinned by the same mechanism? Evidence from pupillometry. *Biological Psychology*, 141, 64–74. <https://doi.org/10.1016/j.biopsycho.2019.01.002>
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44(2), 314–324. <https://doi.org/10.3758/s13428-011-0168-7>
- Näätänen, R. (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. <https://doi.org/10.1017/S0140525X00078407>
- Näätänen, R. (1992). *Attention and brain function*. Lawrence Erlbaum Associates, Inc.
- Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, 42(4), 313–329. [https://doi.org/10.1016/0001-6918\(78\)90006-9](https://doi.org/10.1016/0001-6918(78)90006-9)

- Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1980). Brain potential correlates of voluntary and involuntary attention. *Progress in Brain Research*, *54*, 343–348. [https://doi.org/10.1016/S0079-6123\(08\)61645-3](https://doi.org/10.1016/S0079-6123(08)61645-3)
- Näätänen, R., Kujala, T., & Light, G. (2019). *The Mismatch Negativity: A Window to the Brain* (1st ed.). Oxford University Press. <https://doi.org/10.1093/oso/9780198705079.001.0001>
- Näätänen, R., Kujala, T., & Winkler, I. (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. <https://doi.org/10.1111/j.1469-8986.2010.01114.x>
- Näätänen, R., Paavilainen, P., Tiitinen, H., Jiang, D., & Alho, K. (1993). Attention and mismatch negativity. *Psychophysiology*, *30*(5), 436–450. <https://doi.org/10.1111/j.1469-8986.1993.tb02067.x>
- Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). “Primitive intelligence” in the auditory cortex. *Trends in Neurosciences*, *24*(5), 283–288. [https://doi.org/10.1016/s0166-2236\(00\)01790-2](https://doi.org/10.1016/s0166-2236(00)01790-2)
- Niebuhr, O. (2009). F0-based rhythm effects on the perception of local syllable prominence. *Phonetica*, *66*(1–2), 95–112. <https://doi.org/10.1159/000208933>
- Nieuwenhuis, S., De Geus, E. J., & Aston-Jones, G. (2011). The anatomical and functional relationship between the P3 and autonomic components of the orienting response. *Psychophysiology*, *48*(2), 162–175. <https://doi.org/10.1111/j.1469-8986.2010.01057.x>
- Ou, S., & Guo, Z. (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. <https://doi.org/10.1177/0023830919886604>
- Paavilainen, P. (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. <https://doi.org/10.1016/j.ijpsycho.2013.03.015>
- Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. Oxford University Press.
- Peters, J. (2018). Phonological and semantic aspects of German intonation. *Linguistik Online*, *88*(1). <https://doi.org/10.13092/lo.88.4191>
- Petrone, C., D’Alessandro, D., & Falk, S. (2021). Working memory differences in prosodic imitation. *Journal of Phonetics*, *89*, 101100. <https://doi.org/10.1016/j.wocn.2021.101100>
- Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* [PhD Thesis]. MIT.
- R Core Team, T. (2023). *R: A Language and Environment for Statistical Computing* [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Reeves, C., Schmauder, A. R., & Morris, R. K. (2000). *Stress grouping improves performance on an immediate serial list recall task*. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1638. <https://doi.org/10.1037/02787393.26.6.1638>

- Repp, S. (2020). The Prosody of *Wh* -exclamatives and *Wh* -questions in German: Speech Act Differences, Information Structure, and Sex of Speaker. *Language and Speech*, 63(2), 306–361. <https://doi.org/10.1177/0023830919846147>
- Repp, S., & Seeliger, H. (2020). Prosodic Prominence in Polar Questions and Exclamatives. *Frontiers in Communication*, 5, 53. <https://doi.org/10.3389/fcomm.2020.00053>
- Rietveld, T., & Gussenhoven, C. (1985). On the relation between pitch excursion size and prominence. *Journal of Phonetics*, 13(3), 299–308. [https://doi.org/10.1016/S0095-4470\(19\)30761-2](https://doi.org/10.1016/S0095-4470(19)30761-2)
- Rinne, T., Degerman, A., & Alho, K. (2005). Superior temporal and inferior frontal cortices are activated by infrequent sound duration decrements: An fMRI study. *NeuroImage*, 26(1), 66–72. <https://doi.org/10.1016/j.neuroimage.2005.01.017>
- Rinne, T., Särkkä, A., Degerman, A., Schröger, E., & Alho, K. (2006). Two separate mechanisms underlie auditory change detection and involuntary control of attention. *Brain Research*, 1077(1), 135–143. <https://doi.org/10.1016/j.brainres.2006.01.043>
- Röhr, C. T., Brilmayer, I., Baumann, S., Grice, M., & Schumacher, P. B. (2021). Signal-driven and expectation-driven processing of accent types. *Language, Cognition and Neuroscience*, 36(1), 33–59. <https://doi.org/10.1080/23273798.2020.1779324>
- SanMiguel, I., Corral, M.-J., & Escera, C. (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. <https://doi.org/10.1162/jocn.2008.20078>
- Savino, M., Winter, B., Bosco, A., & Grice, M. (2020). Intonation does aid serial recall after all. *Psychonomic Bulletin & Review*, 27(2), 366–372. <https://doi.org/10.3758/s13423-019-01708-4>
- Seeliger, H., & Repp, S. (2023). Information-structural surprises? Contrast, givenness, and (the lack of) accent shift and deaccentuation in non-assertive speech acts. *Laboratory Phonology*, 14(1). <https://doi.org/10.16995/labphon.6451>
- Seidl, A., & Johnson, E. K. (2006). Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental Science*, 9(6), 565–573. <https://doi.org/10.1111/j.1467-7687.2006.00534.x>
- Sörqvist, P., Marsh, J. E., & Nössl, A. (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. <https://doi.org/10.3758/s13423-013-0419-y>
- Sóskuthy, M. (2017). *Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction* (arXiv:1703.05339). arXiv. <https://doi.org/10.48550/arXiv.1703.05339>
- Spilisbury, G., Stankov, L., & Roberts, R. (1990). The effect of a test's difficulty on its correlation with intelligence. *Personality and Individual Differences*, 11(10), 1069–1077. [https://doi.org/10.1016/0191-8869\(90\)90135-E](https://doi.org/10.1016/0191-8869(90)90135-E)
- Stepanov, A., Kodrič, K. B., & Stateva, P. (2020). The role of working memory in children's ability for prosodic discrimination. *PLoS ONE*, 15(3), e0229857. <https://doi.org/10.1371/journal.pone.0229857>

- Strauch, C., Wang, C.-A., Einhäuser, W., Van der Stigchel, S., & Naber, M. (2022). Pupillometry as an integrated readout of distinct attentional networks. *Trends in Neurosciences*, *45*(8), 635–647. <https://doi.org/10.1016/j.tins.2022.05.003>
- Streefkerk, B. M. (2002). *Prominence. Acoustic and lexical/syntactic correlates* [PhD Thesis, UtrechtLOT]. <https://dare.uva.nl/search?identifier=44dd581d-7c95-442b-96fe-9a6a1e19c1f3>
- Sturges, P. T., & Martin, J. G. (1974). Rhythmic structure in auditory temporal pattern perception and immediate memory. *Journal of Experimental Psychology*, *102*(3), 377. <https://doi.org/10.1037/h0035866>
- Sussman, E., & Winkler, I. (2001). Dynamic sensory updating in the auditory system. *Cognitive Brain Research*, *12*(3), 431–439. [https://doi.org/10.1016/S0926-6410\(01\)00067-2](https://doi.org/10.1016/S0926-6410(01)00067-2)
- Terken, J., & Hermes, D. (2000). The Perception of Prosodic Prominence. In M. Horne (Ed.), *Prosody: Theory and Experiment: Studies Presented to Gösta Bruce* (pp. 89–127). Springer Netherlands. https://doi.org/10.1007/978-94-015-9413-4_5
- Unger, S. M. (1964). Habituation of the vasoconstrictive orienting reaction. *Journal of Experimental Psychology*, *67*(1), 11–18. <https://doi.org/10.1037/h0044510>
- Vachon, F., Hughes, R. W., & Jones, D. M. (2012). Broken expectations: Violation of expectancies, not novelty, captures auditory attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 164–177. <https://doi.org/10.1037/a0025054>
- van der Burght, C. L., & Meyer, A. S. (2025). Working memory capacity predicts sensitivity to prosodic structure. *OSF Preprints*. <https://doi.org/10.31219/osf.io/qny7x>
- van Rij, J., Hendriks, P., van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the Time Course of Pupillometric Data. *Trends in Hearing*, *23*, 2331216519832483. <https://doi.org/10.1177/2331216519832483>
- van Rij, J., Wieling, M., Baayen, R. H., & Rijn, H. van. (2022). *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs* (Version 2.4.1) [Computer software]. <https://cran.r-project.org/web/packages/itsadug/index.html>
- Ventura, C., Grice, M., Savino, M., Kolev, D., Brilmayer, I., & Schumacher, P. B. (2020). Attention allocation in a language with post-focal prominences. *NeuroReport*, *31*(8), 624–628. <https://doi.org/10.1097/WNR.0000000000001453>
- Wang, C.-A., & Munoz, D. P. (2015). A circuit for pupil orienting responses: Implications for cognitive modulation of pupil size. *Current Opinion in Neurobiology*, *33*, 134–140. <https://doi.org/10.1016/j.conb.2015.03.018>
- Wechsler, D. (1987). *WMS-R: Wechsler Memory Scale--Revised: manual*. Psychological Corp.: Harcourt Brace Jovanovich.
- Wetzel, N., Buttelmann, D., Schieler, A., & Widmann, A. (2016). Infant and adult pupil dilation in response to unexpected sounds. *Developmental Psychobiology*, *58*(3), 382–392. <https://doi.org/10.1002/dev.21377>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (2nd ed. 2016). Springer International Publishing: Imprint: Springer. <https://doi.org/10.1007/978-3-319-24277-4>

- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best Practices and Advice for Using Pupillometry to Measure Listening Effort: An Introduction for Those Who Want to Get Started. *Trends in Hearing*, 22. <https://doi.org/10.1177/2331216518800869>
- Winter, B., & Wieling, M. (2016). How to analyze linguistic change using mixed models, Growth Curve Analysis and Generalized Additive Modeling. *Journal of Language Evolution*, 1(1), 7–18. <https://doi.org/10.1093/jole/lzv003>
- Wood, S. (2017). *Generalized Additive Models: An Introduction with R, Second Edition* (2nd ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315370279>
- Wood, S. (2023). *mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation* (Version 1.9-1) [Computer software]. <https://cran.r-project.org/web/packages/mgcv/index.html>
- Wright, R. D., & Ward, L. M. (2008). *Orienting of attention*. Oxford University Press.
- Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The Pupil Dilation Response to Auditory Stimuli: Current State of Knowledge. *Trends in Hearing*, 22. <https://doi.org/10.1177/2331216518777174>

