Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Attention Allocation to Deviants with Intonational Rises and Falls: Evidence from Pupillometry

Permalink

https://escholarship.org/uc/item/68q9c2r8

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors

Lialiou, Maria
Harris, Jesse
Grice, Martine
et al.

Publication Date

2024

Copyright Information

Attention Allocation to Deviants with Intonational Rises and Falls: Evidence from Pupillometry

Maria Lialiou (mlialiou@uni-koeln.de)
IdSL1, Albertus-Magnus-Platz, 50923, Cologne, Germany

Jesse A. Harris (jharris@humnet.ucla.edu)
UCLA, 3125 Campbell Hall, Los Angeles, CA, 90095, USA

Martine Grice (martine.grice@uni-koeln.de)
I*f*L – Linguistik, Herbert-Lewin-Straße 6, 50931 Cologne, Germany

Petra B. Schumacher (petra.schumacher@uni-koeln.de)
IdSL1, Albertus-Magnus-Platz, 50923, Cologne, Germany

Abstract

This pupillometric study investigates the relevance of domain-final intonation for attention-orienting in German, employing a changing-state oddball paradigm with rising, falling and neutral intonation on deviant stimuli. Pupil dilation responses (PDR) to deviants were shown to be affected by their intonation contours, strengthening the case for the role of intonational edge tones in attention-orienting. Moreover, the magnitude and duration of the PDR response was higher for rises than falls, indicating the fundamental role of intonational rises for the activation of the attention-orienting mechanism in speech perception.

Keywords: attention-orienting; intonation; pupillometry; German

Introduction

Humans are confronted everyday with an influx of sounds coming from several sources. In a given auditory environment, some of the sound events might be unexpected, rare, or new. The cognitive system has the ability to detect such sounds, and consequently activate an attention-orienting response. In auditory perception, it is well attested that the orienting response is sensitive to the physical properties of the signal. For example, rises in amplitude or pitch of sine waves form essentially intrinsic warning cues with ripple effects, prompting auditory looming, and, in turn, activating attentional resources, which are expressed in a series of enhanced neural, psychological, or physiological reflexive responses (e.g., Näätänen et al., 1978; Rinne et al., 2005; 2006, Macdonald & Campbell, 2011; among others). In language perception, intonational rises are also pivotal cues, as they are used for getting interlocutors' attention (e.g., for asking questions across languages, e.g., Dingemanse et al., 2013), directing listeners' attention towards the most important part and/or an unexpected change in an utterance (e.g., Lialiou et al., 2024), even guiding attention in serial recall tasks (e.g., Savino et al., 2020 for edge tones; Röhr et al., 2022 for both pitch accents and edge tones). In this pupillometric study, we thus delved into the relevance of domain-final rises (the reflex of phrase-final high edge tones) for attention-orienting in German. Using a *changing-state oddball paradigm*, in which auditory sequences of sequentially ordered ascending numbers (*standards*) were occasionally interspersed with an out-of-sequence number (*deviant*), we investigated whether rising pitch in speech takes on a special role in attention-orienting, by measuring pupil dilation responses (PDR) to arithmetic deviancies.

Over the years, many studies have investigated the mechanism that underpins attention-orienting. One of the mechanisms deemed to underlie auditory attention-orienting is the *expectancy-violation* mechanism (e.g., Vachon et al., 2012; Näätänen et al., 2019 among others). More specifically, the auditory system is able to develop expectations by detecting regularities in the sound environment, and thus predict upcoming sound events. When a deviant sound occurs instead of an anticipated event, it attracts attention, initiating an orienting response. The current study addressed how attention-orienting in response to deviancies in numeric sequences is modulated by the intonation pattern that these deviancies feature. In our study, standard numbers featured a shallow falling intonation (hereafter, *neutral intonation*), a pattern that can be used on non-final items of a list in German. Deviant numbers were produced with one of three intonational patterns: neutral, final rising, or final falling intonation. Although both domain-final rises and falls naturally occur at the end of sequences (or subsequences) in lists, rises tend to indicate that there is more to come, whereas falls indicate finality (e.g., Grabe, 1998; Baumann & Trouvain, 2001; Chen, 2003; Peters, 2018).

Our prediction was that the presentation of a deviant number (as in *23, 24, 25,* ***27****, 28...*) will disrupt the anticipated pattern, evoking a shift in the attentional resources. This prediction is based on the claim that the attention-orienting response is underpinned by an expectancy-violation mechanism (e.g., Vachon et al., 2012). Assuming that attention-orienting is indexed, at least in part, by PDR excursions, the disruption of the anticipated pattern caused by the deviant number will capture attention, which in turn will induce a PDR. Based on the association between attention-orienting and rising pitch found in previous work (e.g.,

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

Lialiou et al., 2024), we further predicted that a final rise on deviants will result in greater disruption, thus inducing a greater PDR compared to deviants with final falls or neutral intonation.

The attention-orienting response has been linked to various physiological responses, such as heart and respiratory rate, electrodermal, vasoconstrictive, neural, as well as pupillary reactions (e.g., Hughes et al., 2007; Liao et al., 2016; Marois et al., 2018; Marois et al., 2019; Näätänen et al., 2019 among others). Pupillary responses, and more specifically, PDR has been mainly used as a correlate of cognitive effort, heightened attention, expectancy violation, and memory consolidation (see Winn et al., 2018 for review). PDR is of particular interest for us because it has been identified as a rigorous (psycho)physiological index of auditory attention-orienting, by a growing number of studies, similar to the neural MMN/P3 responses observed using event-related potential measures. Studies have also reported that PDR is sensitive to the physical characteristics of the deviant sound events. For example, it has been found that the more intense the acoustic properties of a deviant, the larger the magnitude of the PDR (e.g., Liao et al., 2016; Wetzel et al., 2016; Marois et al., 2018). These findings align with (neuro)cognitive and linguistic research on the relevance of different auditory cues on attention-orienting, suggesting that in both cognitive and linguistic domains, the intrinsic properties of deviant events are essential in attracting attention (e.g., Näätänen et al., 1978; Rinne et al., 2006; Paavilainen et al., 2006; Bach et al., 2008; Macdonald & Campbell, 2011; Chobert et al., 2011; Tsang et al., 2011; Dingemanse et al., 2013; Liao et al., 2016; Wetzel et al., 2016; Li & Chen, 2018; Marois et al., 2018; Lialiou et al., 2024; among others).

Methods

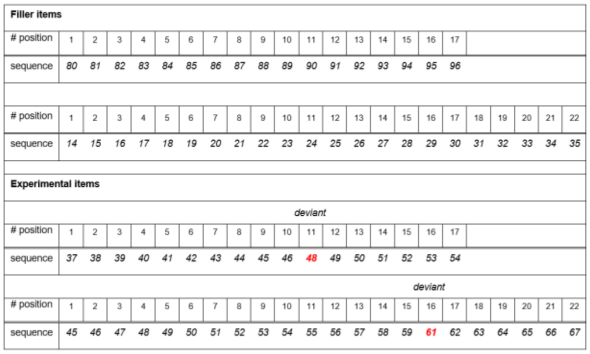
Participants

Sixty native speakers of German (fifty-four female; six male), aged between 19 to 38 years old (mean age = 22 years, and 6 months SD = 3.34) participated in this study. Participants were provided with a written informed consent in accordance with the Declaration of Helsinki and in compliance with the ethics clearance from the Ethics Board of the *Deutsche Gesellschaft für Sprachwissenschaft* (DGfS). Participants received reimbursement for their participation (either course credit or monetary compensation). None of them reported any speech, hearing, or neurological impairment.

Speech Materials

The auditory stimuli comprise sequentially ascending ordered lists of numbers, i.e., seriatim ascending numeric sequences, consisting of 17 numbers (medium sequence length) or 22 numbers (long sequence length). They were combined with three different prosodic realizations (neutral, rise, fall) on the deviant number. In total, 36 different experimental numeric sequences were constructed for this study, that is 6 different numeric sequences x 2 different lengths x 3 prosodic conditions. They were combined with 36 filler items, which did not include a deviant in the sequence. Figure 1 illustrates instantiations of filler and experimental materials.

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



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																						
											deviant											
# 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*					
																deviant						
# 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 illustrating the different lengths of the numeric sequences used as well as the two different deviant positions in the experimental items.

The rise and fall conditions in the prosodic manipulation 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* [ˈfʏnftsɪç]), four (e.g., 52 *zweiundfünfzig* [ˈtsvaɪuntfʏnftsɪç]) or five syllables (e.g., 57 *siebenundfünfzig* [ˈziːbənuntfʏnftsɪç]) always with primary stress on the first syllable, allowing enough time for the different intonation contours to unfold. The deviant numbers consisted mainly of four syllables (32 out of the 36 target numbers), while four of them were pentasyllabic. An example of a numeric sequence for each of the three prosodic conditions is depicted in Figure 2. In the rising condition, all numbers in the sequence were produced with the same intonation, that is a shallow falling contour (hereafter, neutral intonation), an intonational contour that can be featured on non-final items of a sequence or list in German, except for the deviant which was prosodically manipulated, realized with a final rising intonational contour, the reflex of a phrase-final high edge tone. In the falling condition, similarly to the rising one, all numbers in the sequence were produced with the same neutral intonation, but the deviant was prosodically manipulated, this time realized with a final falling intonational contour, the reflex of a phrase-final low edge tone. A neutral condition, in which all numbers in the sequence were produced with the shallow falling intonation without prosodic manipulation of the deviant, served as baseline condition. Across all three conditions, the last number of every sequence was realized with a final falling intonational contour in order to mark the end of that sequence.

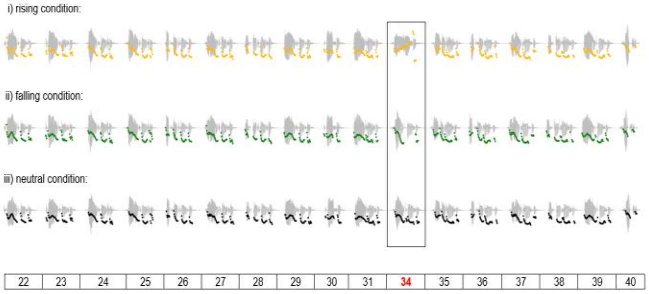


Figure 2: Speech waveform & F0 contour (100-300Hz) of a sample experimental sequence per condition.

The filler items were constructed without a deviant in the sequence, enhancing the creation of the sequentially ascending ordered numeric pattern. The filler items consisted of a different range of numbers compared to the experimental ones to ensure variability in the sequences' construction. This range was between 2 and 99, consisting 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 regards to the prosodic conditions to ensure that participants could not identify the deviant in the experimental items just by tuning into the prosodic manipulation. Twelve of the filler items highlighted a number in the sequence with a boundary rising intonational contour (comparable to the rising condition), another twelve items highlighted a number in the sequence with a boundary falling intonational contour (comparable to the falling condition), and finally, twelve items had no additional prosodic manipulation (comparable to the neutral condition). Fillers differed from experimental items in that the position of the prosodically manipulated number in the sequence was fully randomized to ensure that participants would not be able to predict 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 fully counterbalanced design. Specifically, items were distributed in three lists. Each list presented 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-randomized 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 three 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.

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 a natural speech production of the items, first, the speaker produced all numbers from 0 to 100 in separate blocks as a function of prosodic conditions (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 numerical sequences using Praat (Boersma & Weenink, 2023). The inter-stimulus interval between spliced numbers was 100 ms. The average duration of the medium and long length sequences was 24.86 ms and 25.21 ms respectively. All stimuli used in the experiment were normalized at −23 LUFS but not manipulated further.

Acoustic Characterization of Speech Materials

For the acoustic characterization of the deviant numbers in the sequences, we employed the relative Delta f0 (Δf0) metric from the *ProPer* toolbox, an open-source toolbox based on continuous measurements of periodic energy and f0 (Albert et al., 2020; Albert, 2023). Δf0 describes the f0 trajectory across syllables, using both f0 and periodic energy, indicating f0 changes from syllable to syllable by calculating the difference from the previous one. The raw Δf0 is measured in Hz, yet in this analysis we used relative Δf0 values (relative Δf0 = raw Δf0/speaker's f0 range; for more on Δf0 see Albert, 2023). Δf0 average is centered at 0, thus high/rising syllables exhibit values higher than 0, while low/falling syllables exhibit values lower than 0 (low < 0 < high).

Figure 3 depicts relative Δf0 values per syllable as a function of prosodic condition across quadrisyllabic deviant numbers. In the rising condition (see yellow color), mean Δf0 starts at a mid-level and from first to second syllable, rises shallowly, remains on the same level until syllable three, and then steeply rises towards the last syllable, i.e., the right edge of the word. In the German Tones and Breaks Indices framework (GToBI; Grice et al., 2005), this contour could be described using the label H* ^H-%. The mean Δf0 of both falling and neutral conditions (see green and black colors, respectively)starts a bit higher than the mean Δf0 in the rising condition, and gradually falls from the first to the last syllable. Comparing falling and neutral conditions, the first syllable of falling is a bit higher than neutral. In addition, the drop i) from first to second and ii) from second to third syllables, is steeper for falling than neutral, while iii) the drop towards the last syllable is smaller in falling than in neutral condition. Differences between conditions are depicted in Figure 3.

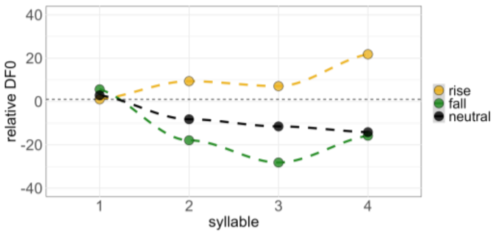


Figure 3: Mean relative $\Delta f0$ values per syllable (x-axis) across prosodic conditions in quadrisyllabic numbers.

In GToBI, the neutral and the falling conditions would fall under the same label (H* L-%), rendering us unable to explain the gradient difference in the meaning between the two contours. As the neutral contour falls less steeply than the falling contour, less clearly marking the end of a unit, it can be expressed on non-final items in a sequence. The falling contour, in contrast, due to its steep fall and extra low pitch towards the end, unambiguously marks the end of a unit, indicating finality of a sequence. Lastly, the rising contour can also mark the end of a unit, yet is functionally different from the falling, in it indicates continuity.

Procedure

After briefing, participants started the experiment. The auditory stimuli were presented using 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 II head-mounted eye-tracker (SR Research Ltd.). Prior to the beginning of the task, the system was calibrated to the participants' dominant eye 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 not informed that the numerical sequences would contain deviants. In order to keep them engaged with the task, participants had to answer a comprehension yes/no question, in 35% of the trials, related to the numeric sequence they heard, by pressing a button on the keyboard.[1] Written instructions were also provided. The experiment started with a practice phase of five items for which they 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 an 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 at the black cross that would appear on the screen. Twenty-five of the trials were followed by a comprehension question which was presented on the screen and participants had to press one of two buttons on the keyboard to indicate their yes/no response. Following the offset of each trial, a grey screen with a black dot was displayed, serving as 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 for the next trial to start. Figure 4 depicts a schematic illustration of an experimental trial.

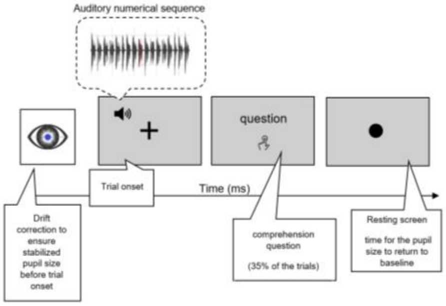


Figure 4: Schematic illustration of an experimental trial.

Data Processing and Statistical Analyses

Data processing and statistical analyses were conducted in R, version 4.1.2 (R Core team, 2021), using the R packages *ggplot2 3.3.5* (Wickham, 2016), *itsadug 2.4.2.* (van Rij et al., 2022), *mgcv 1.8-42* (Wood, 2011), *PupilPre 0.6.2.* (Kyröläinen et al., 2019), and *tidyverse 1.3.1* (Wickham et al., 2019).

Data Pre-processing

Pupil data was exported using SR Research Data Viewer (v.4.3.210), and were further processed using the R-package *PupilPre* (v. 0.6.2; Kyröläinen et al., 2019). Pupil data was re-aligned to 100 ms prior to the onset of the deviant number and continued for 2.900 ms. First, 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 was baseline normalized (by trial) using the average of 100 ms preceding the onset of the deviant. Finally, normalized data was down-sampled to a rate of 100 Hz (i.e., 10 ms time bins).

Inference Criteria

This analysis tests whether deviant numbers produced with rising intonation, due to its attention-orienting function, will induce a stronger PDR compared to when realized with

---

[1] Participants' response accuracy to these questions ranged between 86-100% indicating high engagement with the task.

falling intonation or when its intonation does not differ from that of the standard numbers, i.e., the deviant is produced with neutral intonation.

The PDR was normalized and modelled using Generalized Additive Mixed Modelling (GAMMs), which have been shown to be effective for analysing pupil data, as they account for nonlinear random effects, as well as the inherent autocorrelation of time-course data (see van Rij et al., 2019). In particular, GAMMs are 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, while smooth terms allow for differences in the shape of the curves. GAMMs also account for random effect structures, using random smooths (e.g., Śóskuthy, 2017). Random smooths expand the principle of smooth terms to the random effects, by fitting separate smooths at each value of a grouping variable, allowing thus subjects and/or items to have different curve shapes.

PDR was modelled as a function of the ordered factor[2] 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, and ii) rise vs. neutral. The model included PROSODIC CONDITION both as parametric term, testing for overall height differences in PDR curves between prosodic conditions, and as smooth reference term, capturing the shape effects in the reference level of 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.

The next section presents results reporting model predictions, as well as a plot of the GAMM smooths depicting the effect of each condition on the PDR within the time window of interest. Finally, we used the simultaneous confidence interval test to examine the contrast between fall and neutral prosodic conditions. The simultaneous confidence interval test can be used as a proxy to 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.

Results

Figure 5 illustrates results of the PDR over time as a function of prosodic condition. When comparing rising to falling intonation, rises are differentiated from falls by the *shape* of the PDR curve (smooth difference: EDF = 1.05, F = 3.63, $p = 0.05$), meaning that they exhibit a more sustained effect over time. This result hence suggests that rising deviants lead to a greater disruption of rule-based expectancies, capturing attention for longer compared to falling deviants. When comparing rising to neutral manipulation, PDR to rises is more robust in terms of both the *overall height* (parametric difference: $\beta = -37.05$, $t = -4.38$, $p < 0.001$) as well as the *shape* of the PDR curve (smooth difference: EDF = 3.2, F = 7.04, $p < 0.001$), showing an increased and long-lasting effect, which in turn indicates that attention is oriented towards the rising deviant to a greater extent, and for longer, compared to a neutrally produced deviant. For the contrast between falling and neutral manipulation, 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 ($t = 2.35$), in that its amplitude is larger and prolonged, pointing towards a greater shift of attentional resources towards the falling deviants as compared to neutral ones.

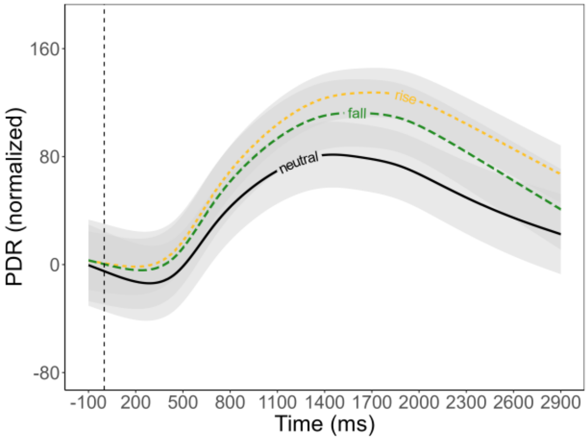


Figure 5: GAMM smooths of PDR (normalized) with 95% confidence intervals, time-locked to the onset of the deviant (zero ms), as indicated by the vertical dashed line. PDR to rising edge tones is illustrated in a dashed orange line, PDR to falling edge tones in a long-dashed purple line, and PDR to neutral intonation in a black solid line.

To briefly summarize the main findings, the two types of edge tones elicit a greater pupillary response (both in magnitude and duration) than the baseline neutral intonation. Within the edge tone conditions, rising intonation resulted in a greater pupillary response than falling intonation, manifesting as a more sustained effect over time. In the next section, we argue that the gradient effects of intonation on PDR response indicate different modulations of attention-orienting in each prosodic case.

---

[2] 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).

Discussion and Conclusion

The current work investigated the relevance of domain-final rises (the reflex of phrase-final high edge tones) for attention-orienting in German, by utilizing a changing-state oddball paradigm in a pupillometry study. Specifically, we investigated whether deviant numbers in sequentially ascending sequences can capture more attention, evoking thus more robust PDRs, when realized with a final rise compared to when they are realized with a final fall or when their intonation does not differ from that of the standards (the neutral case).

To begin with, our results show that deviant numbers do elicit a PDR regardless of the intonation they featured, corroborating the idea that attention-orienting is associated with the expectancy violation mechanism. Now, when the intonation of deviants differed from the intonation of standards (with both rising and falling tones), deviants elicited a stronger PDR than the neutral condition, in which the intonation of the deviant did not differ from the standards, indicating that final edge tones garner additional attentional resources towards the violation. This finding is compatible with results from neurocognitive studies, in which signals with falling acoustic properties were also found to attract some attention, although processed differently than rises (e.g., Macdonald & Campbell, 2011).

Moving to the comparison between rising and falling intonation, we find a subtle shape difference, with PDR to rises lasting longer compared to PDR to falls. Both rises and falls occurred in a context where they could both mark the end of a smaller unit within the sequence. Thus, when rises and falls are contextualized in a similar way, rises attract more attention, as shown by the PDR responses. This finding highlights the attention-orienting function of rising pitch, suggesting that also domain-final rises on deviant stimuli enhance the ability of the deviant to attract attention.

These findings also strengthen the case for a role of edge tones in attention-orienting. Intonational events are phonologically anchored to specific positions in the prosodic structure, that is either they are associated with the stressed syllable, called pitch accents, or with the edges of constituents, called edge tones (in the current case a domain-final edge tone). Traditional theories like the autosegmental-metrical theory (e.g., Ladd, 2008) strictly associated pitch accents with a highlighting function, while edge tones were associated with a phrasing function. In that sense, it has been claimed that accentual rises are better in directing listeners' attention than rises at edges. This has already been called into question by recent work showing that during on-line processing attention is not drawn towards a specific point in the pitch trajectory, but rather that the direction of the pitch contour (rising vs. falling) is primarily relevant for attention-orienting, with rising intonation, taking on a special role, including domain-final rises (Lialiou et al., 2024). Our results contribute to this novel discovery that rises associated with the edges of constituents can also direct listeners' attention. Further, results from serial recall studies (Savino et al., 2020; Röhr et al., 2022) report that rising edge tones marking the final item of non-final triplets boosted recall accuracy of the whole triplet, hence orienting attention to the whole domain. Building on these results, we speculate that domain-final rises attract more attention than falls because rising edge tones attract attention to the entire domain. It is possible that rising edge tones further encourage the listener to more quickly integrate the deviant with previous items in the same domain, thus increasing listeners' awareness of the violation.

Overall, our results support the main claim that, in spoken language, rising intonation takes on a special role in attracting attention. Rising intonation evoked the greatest PDR response, leading to a more pronounced shift of attentional resources compared to the other conditions. Our findings support the idea of an attentional bias towards pitch rises, and extend it from the cognitive domain of audition to the linguistic domain. One could thus argue that the neural architecture of the two domains, cognition and language, share a common pool of mechanisms that are crucial for the attention-orienting network. One of these mechanisms is the signal-based mechanism (also called the *intrinsic auditory mechanism* by Assaneo et al., 2019) for which the signal properties are fundamentally important. The intrinsic properties of the signal are thus essential for both auditory and speech perception, as they feed the signal-based mechanism. In cognition, rising pitch could serve as a particularly robust warning cue, precisely because it is acoustically salient, preparing the nervous system by activating attentional resources (see Bach et al., 2008). As a consequence, listeners allocate more attentional resources in response to unexpected stimuli, initiating an involuntary switch of attention. This switch, in turn, leads to neural, psychological, and physical reflexive responses. In language, rising pitch orients attention towards the most prominent/important information in the speech signal, even when this pitch rise is attributable to edge tones, a process that is crucial for successful interpretation and speech decoding (e.g., Lialiou et al., 2024). This finding is in line with the findings of Lialiou et al. (2024) who show that the phonological status of a pitch event (head or edge associated) is not of primary relevance for attention-orienting. Rather, attention-orienting is driven in large part by the direction of the pitch contour and the appropriateness of the contour in the linguistic context. Here, both rises and falls are appropriate for the list context as they function in a similar way: They mark the end of a smaller sequence unit (subsequence) in a list, including deviants who violate an expected pattern. Thus, *meaningful* rises are taking priority over meaningful falls precisely because of their acoustic prominence.

To conclude, we present evidence from pupillometry that the intrinsic properties of the sensory input, i.e., signal-driven properties, are fundamental for the activation of the attention-orienting mechanism in speech perception. The role of rises is crucial both in cognitive and linguistic domains, as their acoustic properties appears to initiate similar attention-orienting processes across domains.

Acknowledgments

We would like to thank Brita Rietdorf and Claudia Kitler for the great help with recruiting participants and running the experiment. A big thank you goes to Márton Sóskuthy for the invaluable feedback on our GAMM analysis. Last but not least, we thank the SFB 1252 Prominence in Language, project A01, funded by the German Research Council (Project-ID 281511265), for supporting this work.

References

Albert, A. (2023). A model of sonority based on pitch intelligibility. In *Language Science Press*. Language Science Press.

Albert, A., Cangemi, F., Grice, M., & Ellison, T. M. (2020). *ProPer: PROsodic analysis with PERiodic energy* [Computer software]. OSF. https://osf.io/28ea5/

Assaneo, M. F., Rimmele, J. M., Orpella, J., Ripollés, P., de Diego-Balaguer, R., & Poeppel, D. (2019). The Lateralization of Speech-Brain Coupling Is Differentially Modulated by Intrinsic Auditory and Top-Down Mechanisms. *Front. Integr. Neurosci.*, 13, 28.

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.

Baumann, S., & Trouvain, J. (2001). On the prosody of German telephone numbers. *Proceedings of the seventh European Conference on Speech Communication and Technology* (pp. 557–560).

Chen, A. (2003). Language Dependence in Continuation Intonation. *Proceedings of the fifteenth International Congress of Phonetic Sciences* (pp. 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.

Dingemanse, 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.

Grabe, E. (1998). *Comparative Intonational Phonology: English and German.* MPI Series in Psycholinguistics 7, Wageningen, Ponsen en Looien.

Grice, M., Baumann, S., & Benzmüller, R. (2005). German Intonation in Autosegmental-Metrical Phonology. In Sun-Ah Jun (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 55–83). Oxford University Press.

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.

Kyröläinen, A., Porretta, V., van Rij, J., Järvikivi, J. (2019). *PupilPre: Tools for Preprocessing Pupil Size Data*. R package version 0.6.2.

Ladd, D. R. (2008). *Intonational Phonology* (2nd ed.). Cambridge University Press.

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.*

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*, 43.

Li, X., & Chen, Y. (2018). Unattended Processing of Hierarchical Pitch Variations in Spoken Sentences. *Brain and Language, 183,* 21–31.

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.

Marois, A., Labonté, K., Parent, M., & Vachon, F. (2018). Eyes have ears: Indexing the orienting response to sound using pupillometry. *International journal of psychophysiology: official journal of the International Organization of Psychophysiology, 123*, 152–162.

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.

Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica, 42*(4), 313–329.

Näätänen, R., Kujala, T., & Light, G. (2019). The mismatch negativity (MMN): An introduction. In R. Näätänen, T. Kujala, & G. Light (Eds.), *The Mismatch Negativity: A Window to the Brain.* Oxford University Press.

Paavilainen, P., Arajärvi, P., & Takegata, R. (2007). Preattentive detection of nonsalient contingencies between auditory features. *NeuroReport, 18*(2), 159–163.

Peters, J. (2018). Phonological and semantic aspects of German intonation. *Linguistik Online, 88*(1), Article 1.

{R Core Team}. (2021). *R: A Language and Environment for Statistical Computing* [Computer software]. R Foundation for Statistical Computing.

Rinne, T., Degerman, A., & Alho, K. (2005). Superior temporal and inferior frontal cortices are activated by infrequent sound duration decrements: An fMRIstudy. *Neuro Image, 26*(1), 66–72.

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.

Röhr, C. T., Savino, M., & Grice, M. (2022). The effect of intonational rises on serial recall in German. *Proceedings of the eleventh International Conference on Speech Prosody* (pp. 759–763).

Savino, M., Winter, B., Bosco, A., & Grice, M. (2020). Intonation does aid serial recall after all. *Psychonomic Bulletin & Review, 27*(2), 366–372.

Sokolov, E. N. (1963). Higher Nervous Functions: The Orienting Reflex. *Annual Review of Physiology, 25*(1), 545–580.

Sóskuthy, M. (2017). Generalised additive mixed models for dynamic analysis in linguistics: a practical introduction. *arXiv e-prints*, arXiv-1703.

Tsang, Y. K., Jia, S., Huang, J., & Chen, H. C. (2011). ERP Correlates of Pre-Attentive Processing of Cantonese Lexical Tones: The Effects of Pitch Contour and Pitch Height. *Neuroscience Letters, 487*(3), 268–72.

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.

van Rij, J., Wieling, M., Baayen, R., & van Rijn, H. (2022). *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs*. R package version 2.4.1.

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.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (2nd ed. 2016). Springer International Publishing : Imprint: Springer.

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.

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*, 2331216518800869.

Wood, S.N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B), 73*(1), 3-36.