Auditory Processing of Intonational Rises and Falls in German: Rises Are Special in Attention Orienting

Maria Lialiou, Martine Grice, Christine T. Röhr, and Petra B. Schumacher

Abstract

■ This article investigates the processing of intonational rises and falls when presented unexpectedly in a stream of repetitive auditory stimuli. It examines the neurophysiological correlates (ERPs) of attention to these unexpected stimuli through the use of an oddball paradigm where sequences of repetitive stimuli are occasionally interspersed with a deviant stimulus, allowing for elicitation of an MMN. Whereas previous oddball studies on attention toward unexpected sounds involving pitch rises were conducted on nonlinguistic stimuli, the present study uses as stimuli lexical items in German with naturalistic intonation contours. Results indicate that rising intonation plays a special role in attention orienting at a pre-attentive processing stage, whereas contextual meaning (here a list of items) is essential for activating attentional resources at a conscious processing

stage. This is reflected in the activation of distinct brain responses: Rising intonation evokes the largest MMN, whereas falling intonation elicits a less pronounced MMN followed by a P3 (reflecting a conscious processing stage). Subsequently, we also find a complex interplay between the phonological status (i.e., accent/head marking vs. boundary/edge marking) and the direction of pitch change in their contribution to attention orienting: Attention is not oriented necessarily toward a specific position in prosodic structure (head or edge). Rather, we find that the intonation contour itself and the appropriateness of the contour in the linguistic context are the primary cues to two core mechanisms of attention orienting, pre-attentive and conscious orientation respectively, whereas the phonological status of the pitch event plays only a supplementary role. ■

INTRODUCTION

Voluntary attention shields the processing of currently important information from irrelevant information. Unexpected sound changes outside of the current attentional focus may break through this shield causing an involuntary switch of the attentional resources toward these unexpected changes (for reviews, see Näätänen, Kujala, & Light, 2019). It has been claimed that one of the auditory cortex change detection mechanisms that triggers involuntary attention is reflected in the neurophysiological response called MMN (e.g., Näätänen, 1992). The present study is concerned with the processing of intonational rises and falls in German when presented unexpectedly in a stream of repetitive auditory stimuli. In an EEG study, using an auditory passive oddball paradigm, we investigate whether rising pitch in speech is special in attention orienting by measuring MMN responses to intonational changes.

Intonation is a fundamental characteristic of spoken language and is usually described as the melody of speech or tone of voice. When we speak, intonation conveys meaning over and above the meaning of the words and sentences bearing it. For example, the intonation of an utterance can differentiate between a number of different

functions. In particular, an intonational rise can indicate that an utterance is incomplete and that there is more to come, rather than being complete, or that it is a question rather than a statement. Speakers also use intonation to highlight particular information in their utterances or to chunk information into smaller units. Both of these functions, namely, highlighting and chunking, serve in orienting listeners' attention to informative parts of utterances. Prosodic highlighting and chunking find expression in phonological choices, such as pitch accents (i.e., tonal movements associated with the stresses syllable [head]) and edge tones (i.e., tonal movements associated with the edge/boundary of a prosodic constituent) as well as in acoustic parameters such as scaling and timing of fundamental frequency (f0), intensity, segmental durations, and spectral characteristics. It is generally assumed that pitch accents serve to highlight words and constituents, whereas edge tones serve to chunk utterances into smaller units.

In auditory processing, the physical properties of the acoustic signal are essential for attracting attention. For instance, rises in amplitude of sine waves prompt auditory looming effects, indicating that the sound source is approaching. They thus serve as warning cues that activate attentional resources (e.g., Bach et al., 2008). Falling acoustic signals may also attract attention, but it has been claimed that they are experienced mostly as fading

University of Cologne

(indicating a receding sound source), and thus processed differently than rises (e.g., Macdonald & Campbell, 2011). Likewise, in speech processing, intonational rises are pivotal cues for prominence and attention. Rising pitch accents, marking new topics or referents, and different types of focus, especially contrastive or corrective focus (e.g., Lorenzen, Roessig, & Baumann, 2022; Grice, Ritter, Niemann, & Roettger, 2017; Baumann & Schumacher, 2012; Bolinger, 1985, 1989), allocate attention to the words that bear them, facilitating their processing. Rising pitch accents, when used as a focus cueing device, also guide attention toward semantic incongruencies, because they lead to more elaborate processing of the focused information (Ventura et al., 2020; Wang, Bastiaansen, Yang, & Hagoort, 2011). Moreover, both rising pitch accents and rising edge tones guide attention in serial recall tasks (e.g., Savino, Winter, Bosco, & Grice, 2020, for edge tones; Röhr, Savino, & Grice, 2022, for both pitch accents and edge tones).

In this article, we revisit the idea of an attentional bias (essentially serving as a warning cue) toward sounds with rising as opposed to falling acoustic properties and extend it from general cognition to language. We are particularly interested in the role of intonational rises in attention orienting, that is, whether rising pitch attracts more attention than falling pitch, not only at a general cognitive level but also in linguistic terms. In addition, we investigate the role of the phonological status of the rise (pitch accent or edge tone) in attention orienting, that is, whether only rises on pitch accents play a role or whether edge tones have a similar effect (as in the serial recall results cited above). In an EEG study using an auditory passive oddball paradigm, we measure MMN and the subsequent positivity (P3) responses to unexpected intonational changes, as these brain responses have been claimed to be the neurophysiological underpinnings of involuntary and voluntary attentional mechanisms, respectively (e.g., Näätänen, Kujala, & Winkler, 2011).

Attention Orienting toward Unexpected Sound Events

Attention orienting has been defined as the "alignment of attention with a source of sensory input or an internal semantic structure stored in memory" (Posner, 1980, p. 4). Posner (1980) differentiates between voluntary and involuntary attention orienting in that the former (a detecting cognitive act) refers to conscious processing of the input, whereas the latter allows the listener to link responses to the input before it has been consciously processed. A large body of research has been concerned with the investigation of attention orienting toward unexpected sound changes, with the function of the auditory processing system in focus.

Our auditory processing system entails the ability to predict prospective sound events by detecting regularities in the sound environment (Winkler, 2007; Näätänen &

Winkler, 1999). Violations of the anticipated events generate an electrophysiological response in the brain, the MMN. MMN is defined as a regularity-violation response elicited by any perceptible change in the auditory stimulation (Näätänen et al., 2019), even during unattended stimulation (for reviews, see Näätänen et al., 2019; Näätänen, Paavilainen, Rinne, & Alho, 2007). MMN is claimed to be an automatic, pre-attentive response activating an involuntary attention switch toward the unexpected auditory change (e.g., Näätänen 2001; Näätänen et al., 2019; Näätänen, 1990, 1992, among others). The MMN is a negative potential with a fronto-central maximum and a peak latency between 150 and 250 msec after the onset of a deviant stimulus, although its latency may vary with the degree of deviation. There are other transient brain responses that indicate some kind of change or deviance detection (e.g., P3a/b, N400, P600), although their elicitation requires some level of saliency or directed attention as can be, for instance, triggered by task demands (Sussman et al., 2014).

Studies have shown that MMN activation is sometimes followed by a positive deflection. This positivity has been identified with the P3 family (e.g., Polich, 2007). Especially when the positivity emerges after the MMN, it has been referred to as the MMN-P3 complex (for a review, see Näätänen, 1992). The P3 response has been reported to reflect conscious processes of novel or salient events (e.g., Duncan et al., 2009; Polich, 1986; Donchin, 1981). Thus, one could argue that MMN and P3 responses reflect the route from pre-attentive to conscious perception of an unexpected auditory change or deviance (Hsu, Tu, Chen, & Liu, 2023; Näätänen et al., 2011). That is, the MMN generator automatically orients involuntary attention toward an unexpected sound event, which could potentially activate the mechanisms of conscious (i.e., voluntary) processing (reflected in the presence of a P3 response) by which this change is brought into awareness.

The Role of Rises in Attention Orienting

Neurocognitive studies have been concerned with the question of how attention orienting toward unexpected sound events is conditioned by different auditory cues. Such studies manipulated the direction of sound changes by modifying a variety of acoustic features such as f0 (perceived as pitch), duration (perceived as length), and intensity (perceived as relative loudness; e.g., Chobert et al., 2012; Macdonald & Campbell, 2011; Paavilainen et al., 2007; Rinne, Särkkä, Degerman, Schröger, & Alho, 2006; Rinne, Degerman, & Alho, 2005; Alain, Woods, & Ogawa, 1994; Näätänen, Gaillard, & Mäntysalo, 1978, among others), reporting an attentional bias toward unexpected sounds with rising as opposed to falling acoustic properties, indexed by a greater MMN amplitude, or elicitation of an MMN-P3 complex (for a review, see Näätänen et al., 2019). Hence, it has been proposed that rising acoustic properties form intrinsic warning cues that

activate attentional resources eliciting automated motor actions or appropriate adaptive responses when needed (e.g., Bach et al., 2008).

Although linguistic research has shown that rising acoustic properties are also pivotal in spoken linguistic communication, previous research on the processing of unexpected auditory events has focused on pure tones only. To our knowledge, only the study of Hsu, Evans, and Lee (2015) has explored the sensitivity of pitch change direction in relation to speech. However, although some of Hsu and colleague's stimuli were produced with a human voice, they were not designed to convey any linguistic meaning. Specifically, the authors tested pitch changes produced on the phoneme [ɑ] with the Mandarin level tone, starting from habitual pitch and increasing f0 for as long as modal phonation was possible. For a second set of stimuli, they synthetically elevated the pitch above the speaker's normal range. Pitch changes in nonspeech stimuli (i.e., pure tones) at the same frequencies were also included in the study. Using an oddball paradigm, Hsu and colleagues investigated whether unexpected small and large rises in spoken pitch attract listeners' attention to a greater extent than unexpected small and large falling pitch; in addition, they asked whether brain responses to spoken pitch rises are different from similar rising pitch changes in pure tones. They found that whereas MMN to changes at normal and synthetically elevated spoken pitch height did not differ as a function of pitch direction (falling vs. rising), it did differ as a function of the size of the change (small vs. large): Large pitch changes evoked a greater MMN. MMN to pure tones equivalent to the speaker's normal pitch height was evoked only by large falling changes in pitch, whereas at elevated levels, it was evoked by small and large pitch rises, and large pitch falls. In addition, P3 was sensitive to the direction of the change, because only rising pitch changes (both small and large) at a normal spoken pitch height evoked a P3. With pure tones, large pitch changes at an elevated pitch height also gave rise to a P3, which, however, was not sensitive to pitch direction. On the basis of the abovementioned results and, specifically, the P3 sensitivity to rising pitch changes in speech, the authors suggested that sudden pitch rises in speech demand more attentional resources than sudden falls, because their presence activates additional conscious processing mechanisms. Crucially, the rise or fall for the stimuli used in Hsu and colleagues involved a change in pitch from one stimulus to the next (i.e., across stimuli), whereas in our experiment, the rise or fall is within the stimulus itself.

Linguistic Rises and Attention

According to prosodic typology and autosegmental-metrical theory (hereafter, AM), there are two basic ways tonal events can be associated with positions in prosodic structure (e.g., Jun, 2014; Ladd, 2008). Pitch accents and edge tones (hereafter, we use the term "edge/boundary tone" interchangeably) have distinct association properties. Whereas pitch accents are associated with stressed syllables, edge tones are associated with initial or final boundaries of smaller or larger constituents. However, functional properties also play a role in this categorization. In West Germanic languages, such as English and German, which make use of both categories (pitch accents and edge tones), pitch accents not only associate with stressed syllables but are also considered to primarily have a highlighting function, cueing prominence. By contrast, edge tones not only associate with initial or final edges of smaller or larger constituents but it is claimed that they are mainly used for phrasing and only secondarily for highlighting. Thus, it appears that association properties of tones (accent/boundary) are prepackaged with distinct functions, as discussed in Grice (2022): Whereas pitch accents cue prominence, boundary tones cue phrasing. It is therefore proposed that accentual rises constitute a better cue in directing listeners' attention than rises at prosodic boundaries. Nevertheless, studies investigating prominence perception and processing call into question the strict dichotomy proposed by the AM theory. Specifically, it has been claimed that prosodic phrasing actually can affect prominence perception (for a discussion, see Grice & Kügler, 2021). Some evidence that rising boundary tones can also cue prominence comes from serial recall tasks of nine-digit sequences in Italian (Savino et al., 2020) and German (Röhr et al., 2022; Grice et al., submitted). These studies found that boundary rises marking the last item of nonfinal triplets facilitated the recall accuracy not only of the digits at the boundary positions but also of the whole group of items. These results reveal that edge tone rises appear to cue prominence on the whole domain they are delimiting. Other evidence from edge tones cueing prominence comes from processing studies focusing on the domain of the word, showing that rising boundary tones facilitate, for example, word segmentation (e.g., Ou & Guo, 2021) or word recognition (Kember, Choi, Yu, & Cutler, 2021).

Therefore, regardless of the functions with which pitch accents and boundary tones are predefined within the AM theory, the prominence patterns in a language can be affected or modulated by different structural positions (for the notion of structural prominence, see Grice & Kügler, 2021; Cangemi & Baumann, 2020; von Heusinger & Schumacher, 2019; Himmelmann & Primus, 2015; Streefkerk, 2002). Put differently, prominence in West Germanic languages is concerned to be expressed in the form of pitch accents, directing listeners' attention to the highlighted information. The stressed syllable, being the docking site for the pitch accent, is the head of the word and therefore occupies an essential position in the linguistic structure. Yet, prosodic boundaries appear to also guide attention toward important elements. This could potentially be because flagging prominent information at privileged positions, such as in the beginning or at the end of an utterance, is crucial for speech processing and planning

(e.g., Ou & Guo, 2021; Seidl & Johnson, 2006). Hence, one could argue that in a complex speech signal such as language, prominence might not necessarily be encoded by cues at one specific structural position, but rather by the combination of cues at different privileged positions.

Signal- and Context-driven Expectations

The processing of information is highly affected by listeners' expectations (e.g., Roessig, Mücke, & Grice, 2019; Friston, 2018; Grice et al., 2017; Huettig, 2015; Clark, 2013). These expectations are in part driven by pure acoustic properties but can also be driven by context- and/or language-specific expectations. For example, a particular accent on a constituent or an inserted pause can create expectations as to the upcoming information. The neural processing of linguistically meaningful pitch variations has been studied at both the lexical- and postlexical level, showing that brain responses are not only activated by the acoustic contrasts in the signal but also sensitive to the timing of the acoustic cues (e.g., Li & Chen, 2018; Tsang, Jia, Huang, & Chen, 2011).

Although previous research on acoustic mismatch detection has targeted signal-based expectations, context may also shape attentional orientation. For instance, Röhr and colleagues (2021) investigated in two EEG studies the role of signal- and expectation-driven effects of prosodic prominence. On the one hand, four different German accentual contours were tested in isolated sentences (steep rise, shallow rise, [steep] fall, and no accent) making the acoustic signal the only source for attention orienting. On the other hand, the most prominent steep rising accent and the less prominent falling accent were tested with regard to expectations as to how exciting/unusual the content of an utterance is by relating the stimuli to an exciting/unusual and neutral (negligible/ordinary) precontext. Results in general indicated that attentional cues, both signal-driven and context-driven, engender positivities of varying latency: that is, (i) a prominent rise on the stressed syllable (and not a rise elsewhere) consumes attentional resources at an early processing stage (reflected in an Early Positivity) and (ii) highlighting induced by the context (i.e., the exciting context) consumes attentional resources at a later processing stage (reflected in a Late Positivity). Moreover, results showed that prior context builds up expectations for upcoming prosodic input, reflected in N400 prediction errors engendered by acoustically unexpected (here prominent) accents as well as contextually inappropriate prosodic realizations. Hence, these studies suggest that attentional orientation and predictive processing reflect discrete stages in the construction of a mental representation during real-time comprehension. Another example of the interaction between signal- and context-driven processing of pitch variation is the study by Liu, Chen, and Schiller (2016). Liu and colleagues investigated the attentive processing of tone and intonation in Mandarin and found that conscious

attentional neural responses (i.e., P300 response) were modulated by the pragmatic context (question vs. statement) in which Tone4 appeared.

In addition, it has been shown that the processing and perception of prosodic prominence can also be shaped by language-specific expectations, such that attention is allocated to different information, based on the language-specific prosodic structure (e.g., Ventura et al., 2020; Chandrasekaran, Krishnan, & Gandour, 2009). Therefore, prior linguistic or discourse context, or even recent speech experience can potentially overwrite the typical acoustic cues that signal prominence (e.g., Kakouros, Salminen, & Räsänen, 2018; Bishop, 2013; Cole, Mo, & Hasegawa-Johnson, 2010) and lead to different processes.

Motivation for the Current Study

In the context of previous neurocognitive studies suggesting that, in a stream of repetitive nonlinguistic auditory events, unexpected sound events with rising acoustic properties attract more attention than sound events with falling acoustic properties, the current study explores the role of rises and falls in pitch attributable to accents and boundary tones in attention orienting, using speech stimuli that convey linguistic meaning. Extending previous work on simple sine waves, we introduce rising and falling intonation on sequentially presented lexical items, which potentially gives rise to a list context with its language-specific expectations.

We recorded listeners' EEG and measured event-related brain potentials related to unexpected auditory deviances in a repetitive auditory stream. Of particular interest in this study are the MMN and the subsequent positivity (P3), which can potentially follow the MMN time window. This is because previous studies have shown that these two brain responses appear to index the path from preattentive to conscious processes of an unexpected auditory change. MMN can be recorded in different stimulus paradigms based on the different aspects of the central auditory processing that one might want to study. One of the most frequently used paradigms is the classic oddball paradigm in passive recordings, which we also utilize in the current study (e.g., Näätänen et al., 2019). In this paradigm, participants are presented auditorily with sequences of repetitive sounds, called standards, occasionally interspersed with a rare sound, called deviant, while they are watching a film with no sound. Thus, it allows for elicitation of brain responses in the absence of attentive listening and can shed light on the nature of the underlying neural mechanisms.

The main question we put forward in this study is whether linguistic rises attract more attention than falls. We test two central hypotheses. Following the auditory looming literature (e.g., Macdonald & Campbell, 2011; Bach et al., 2008) our Hypothesis 1 is that unexpected rises in a sequence of repetitive falls should attract more attention than unexpected falls in a stream of repetitive rises,

evoking thus a more pronounced MMN, potentially followed by a P3 response. Furthermore, given our interest in the processing of pitch rises and falls in both *accentual* and *boundary* positions, we subsequently ask whether the phonological status of the rise and fall affects the underlying processing. In other words, do accentual contours attract more attention than boundary contours or vice versa? Hypothesis 2 is that the extent of the effect may vary as a function of the phonological association (pitch accent vs. boundary tone) of the tonal event (e.g., Ladd, 2008). If the standard AM account holds, postulating *pitch accents as the primary markers of prominence*, we expect a greater MMN effect for accentual over boundary contours.

METHODS

Employing the classic oddball paradigm in passive recordings and using the speech material described in the section Speech Materials, below, we designed four different oddball conditions in which rising and falling f0 contours alternated as standard/deviant sounds. In two conditions standard/deviant sounds are composed of accentual contours (condition1: standard accentual rise/deviant accentual fall; condition2: standard accentual fall/deviant accentual rise), whereas in the other two conditions, standard/deviant sounds are composed of boundary contours (condition3: standard boundary rise/deviant boundary fall; condition4: standard boundary fall/deviant boundary rise). While listening to the stimuli, participants were watching a nature documentary film with no sound (Deep Blue; Fothergill & Byatt, 2003).

Participants

ERP data from thirty-two right-handed participants were recorded for this study. The sample size was determined on the basis of previous studies utilizing the MMN paradigm. Handedness was assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). All participants were monolingual native speakers of German (28 women, 4 men), aged between 19 and 33 years old (mean age = 24.5 years, *SD* = 3.5). 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*. 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 used in the oddball paradigm comprise rising and falling f0 contours realized either on the stressed syllable (accentual contours) or at the boundary, that is, on the last syllable (boundary contours), of four different lexical items (*Banane* "banana," *Limone* "lime,"

Marone "chestnut," *Melone* "watermelon"), resulting in 16 tokens (4 contours × 4 lexical items). For segmental comparability of the lexical items, we selected trisyllabic common German nouns, with simple segmental structure (CV.′CV.CV) and primary lexical stress on the penultimate syllable, mainly composed of voiced sounds to enable a continuous f0 trajectory. In addition, to control for potential word frequency effects on prominence perception (cf. Cole et al., 2010), all items were of approximately the same frequency class.[1]

All stimuli were produced by a phonetically trained female native speaker of German and recorded with a sampling rate of 44100 Hz and 16-bit resolution (mono). To ensure natural speech production of the items, the speaker was asked to produce spontaneously all items in isolation with all intonational contours. To circumvent inconsistencies in the realization of the same intonation across the different items, we selected the most natural sounding item in terms of speech rate, duration, pitch range, and meaning (e.g., a falling contour marking finality, a rising contour marking continuity) for each contour and presented them repeatedly to the speaker as a prompt for the production of the same contour type across lexical items. The speaker repeated the prompt with as little delay as possible, resulting in a natural but very consistent production of the different contours across items as our acoustic analysis indicated. The original production of all stimuli was used in the experiment, normalized at −23 LUFS. Figure 1 illustrates the mean f0 contours as well as the individual f0 contours of all four items superimposed on each other. The figures indicate that the f0 contours were produced in a very consistent manner across items.

Acoustic Characterization of Speech Material

For the acoustic analysis of our stimuli, we used the novel *ProPer* toolbox, an open source toolbox for acoustic analysis of prosodic-related phenomena based on continuous measurements of periodic energy and f0 (Albert, 2023; Albert, cangemi, Grice, & Ellison, 2023). To acoustically describe our stimuli, we used the relative periodic energy mass (henceforth, mass) and the relative Delta f0 (henceforth, Δf0) metrics. All analyses were conducted on the basis of syllabic units. Scripts and data tables of the current analysis are available on the Open Science Framework (OSF) platform (https://osf.io/57ztj/).

Mass is the area under the periodic energy curve between two syllabic boundaries. It is calculated as the integral of duration and power, accounting for these two cues together in one variable, capturing thus the overall prosodic strength of the corresponding syllable (for more discussion on the components of mass and its improved acoustic characterization of prominence see Albert, 2023). Here, we use the relative mass metric that indicates the prosodic strength of one syllable relative to the other syllables in the word. Relative mass values are calculated as the area under the periodic energy curve of the entire
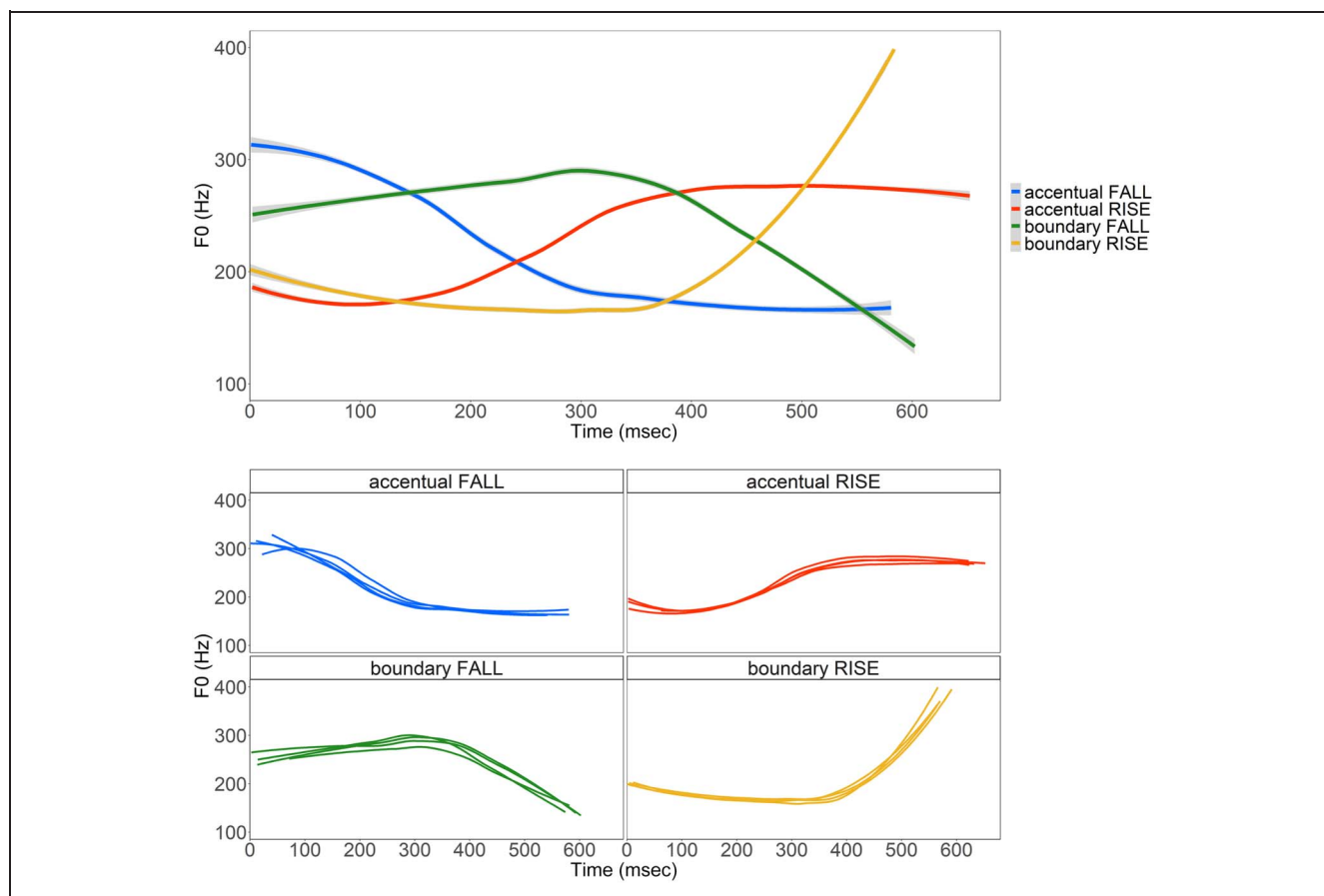
Figure 1. Top: mean f0 contours across items; the vertical dashed line illustrates word onset. Bottom: individual f0 contours of all items, superimposed on each other for each contour.

word divided by the number of the syllables in the word (relative mass = mass of the word/n of syllables). The average mass is centered around one; thus, weak syllables exhibit values lower than one, whereas strong syllables exhibit values higher than one (weak < 1 < strong).

Δf0 describes the f0 trajectory across syllables, using both f0 and periodic energy. The Δf0, first, measures the f0 at the center of mass within each syllabic interval and then computes the difference in f0 between two subsequent syllables. Δf0 thus indicates the f0 change from syllable to syllable by calculating the difference from the previous one. For the first syllable, Δf0 is calculated relatively to speaker's f0 median. The raw Δf0 is measured in Hz, yet in our analysis, we use the relative Δf0 values (relative Δf0 = raw Δf0/speaker's f0 range) presented in percentages (for more discussion on Δf0, see Albert, 2023).

Figure 2 illustrates a representative example of a *periogram*, which is a visual representation of the standard f0 curve of the item *Melone* realized with accentual rising intonation enriched with periodic energy, such that the thicker the line, the more periodic energy. The lower part of the figure depicts the periodic energy curve that modulates the f0 curve in the upper part of the figure. Mass and Δf0 values are illustrated on the same figure.

The consistent production of the f0 contours across items is also shown in the ProPer analysis. Figure 3 depicts

relative Δf0 values per syllable as a function of contour across items as well as per item. In accentual falling contours (see depiction in blue in top images), consistently across items, Δf0 starts H(igh) on the first syllable, falls to a L(ow) on the stressed syllable, and levels (or slightly rises) toward the last syllable in the word. In accentual rising contours (see depiction in red in top images), Δf0 starts L on the first syllable, rises to a H on the subsequent stressed syllable, and levels toward the last syllable. In both falling and rising boundary contours (see bottom images), Δf0 remains on the same level from the first to the subsequent stressed syllable, and then in the former contour (depiction in green), it falls to a L toward the last syllable (boundary) of the word, whereas in the latter (depiction in yellow), it rises to a H.

Let us now move on to the relative mass. Figure 4 presents relative mass values per syllable of each item across accentual and boundary contours. Visual inspection of mass in both accentual and boundary contours shows a slight variability on the values across items. This is expected considering the different phonetic makeup of each syllable across the different lexical items. Despite these slight differences, in the accentual contours, the second syllable, which is stressed and accented, exhibits the greatest mass across items, indicating that it is prosodically the strongest syllable in the word, whereas its overall

Figure 2. A representative example of a periogram. In the upper half of the figure, the blue line illustrates the f0 curve, whereas the red line in the lower half part of the figure depicts the periodic energy curve. The frequency values on the *y* axis correspond only to the f0 curve. The red vertical dashed lines indicate the position of the center of mass (CoM) within syllabic intervals. Mass values on the relative scale are given in numbers below each interval. The short vertical blue lines on top of the f0 curve indicate the center of gravity (CoG), with Δf0 values depicted superimposed on the f0 curve.
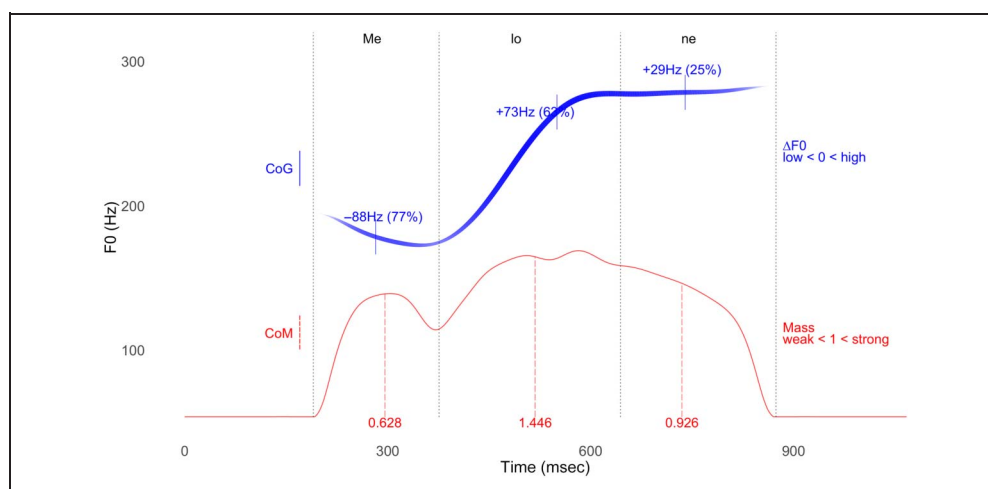

prosodic strength is similar across rising/falling contours. In the boundary contours, the prosodic strength between the stressed (second) and the last syllable that carries the boundary tone do not differ much. One could observe that mass on the stressed syllable is slightly higher in boundary falls than in boundary rises, whereas mass on the last syllable is subtly higher in boundary rises than in boundary falls.
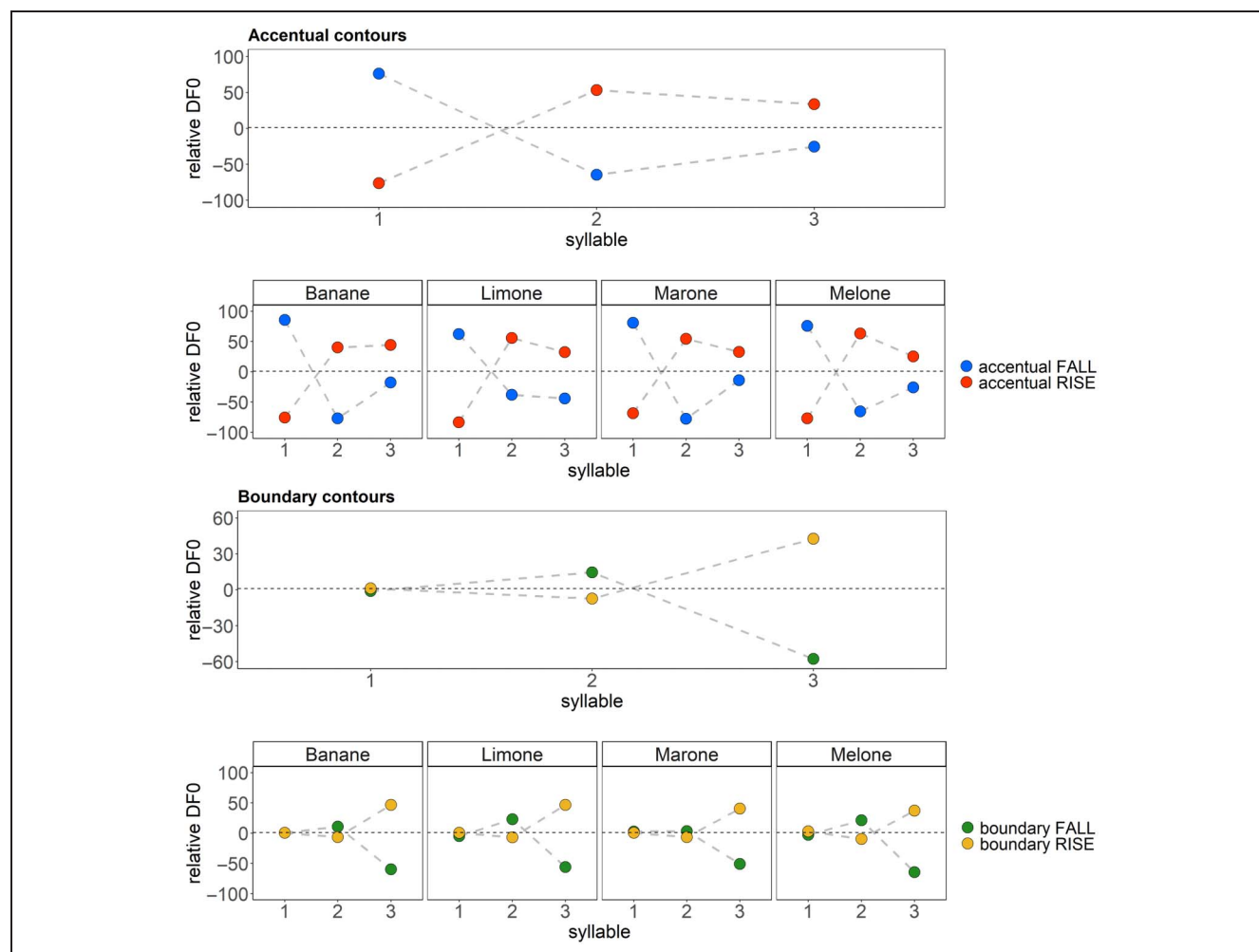
Figure 3. Relative Δf0 values per syllable (*x* axis) across accentual (top images: accentual falls in blue, accentual rises in red) and boundary (bottom images: boundary falls in green, boundary rises in yellow) contours and per item.
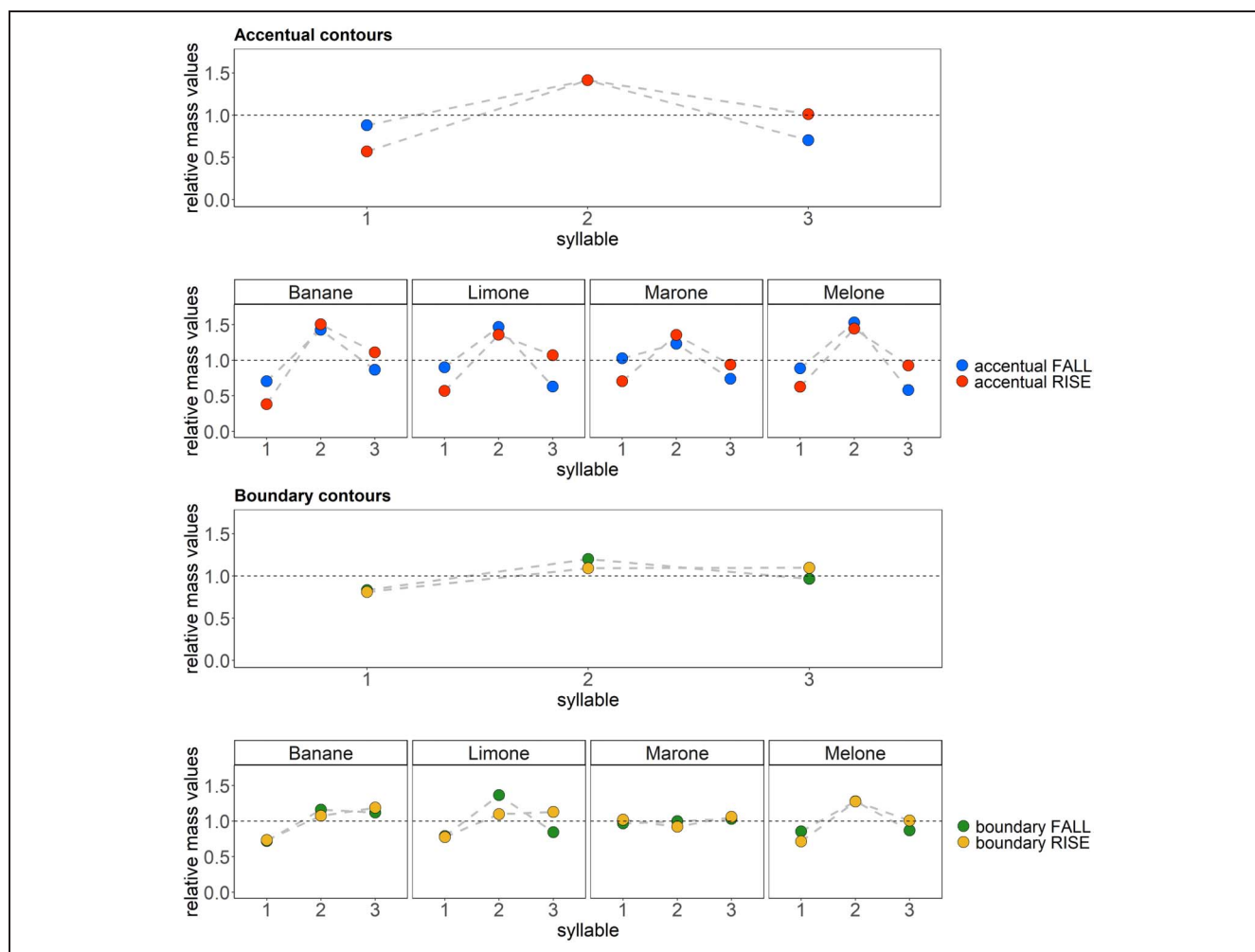
Figure 4. Relative mass values per syllable (*x* axis) as a function of item and intonational contour. Top images illustrate mass values in accentual (rises in red, falls in blue) contours. Bottom images present mass values in boundary (rises in yellow, falls in green) contours.

Table 1 presents means and standard deviations for (i) mass and (ii) Δf0 values per syllable across items for each contour as well as (iii) the total duration of items.

EEG Recording

The EEG was recorded from 32 Ag/AgCl electrodes, amplified with the Brain Vision amplifier, and digitized at a sampling rate of 1000 Hz. The electrodes were mounted in an elastic EEG cap (EasyCap, EasyCap GmbH) and placed on the scalp according to the standard International 10–20 system. The electrical contact between scalp and electrodes was achieved by applying a conductible electrolyte. As the MMN is well documented to have a frontocentral topography (Näätänen et al., 2007), we selected a distribution of mostly frontal electrodes (AF3/4/7/8, F3/4/7/8, Fz, FC1/2/5/6, FCz, C3/4, Cz, CP1/2, CPz, P3/4/7/8, Pz, POz, Oz). The AFz electrode position served as the ground electrode, and additional electrodes were placed to the left and right mastoids for referencing (left) and rereferencing (right) of the EEG channels. To control for eye-movement artifacts, we further recorded the EOG with electrodes

placed to the left and right mastoids at the level of the external canthus of each eye, as well as to the supra- and infra-orbital foramens of the right eye. Electrode impedances were kept below 3 kΩ.

Procedure

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

Participants were presented with standard/deviant f0 contours realized on the same item within condition, but the items across conditions always differed (Latin Square Design: 4 items × 4 oddball conditions). To control for systematic order and frequency effects potentially induced by the exposure to oddball condition and/or

Table 1. Mean and Standard Deviation Values (in Brackets) for Mass and Δf0 Values per Syllable across Items for Each Contour as well as for the Total Duration of Items

Measurements		Accentual Fall	Accentual Rise	Boundary Fall	Boundary Rise
Relative Δf0 (%)	[σ	76.20 (9.99) 113%	−76.13 (6.05) 84%	−1.10 (3.5) 1.5%	1.20 (1.2) 2%
	σ'	−64.50 (18.5) 95%	53.50 (9.5) 60%	14.50 (9.2) 19%	−7.40 (1.5) 15%
	σ]	−25.50 (13.5) 38%	33.70 (7.8) 37%	−57.80 (5.8) 76%	42.70 (4.7) 85%
Relative mass	[σ	0.88 (0.13)	0.57 (0.14)	0.83 (0.10)	0.81 (0.14)
	σ'	1.42 (0.13)	1.42 (0.07)	1.20 (0.16)	1.09 (0.15)
	σ]	0.70 (0.13)	1.01 (0.09)	0.97 (0.13)	1.10 (0.08)
Duration (msec)	item	0.600 (0.01)	0.680 (0.02)	0.630 (0.02)	0.620 (0.02)

item order, we created 16 fully counterbalanced lists with different oddball condition order and item so that each list presented all items and oddball conditions but never the same item across conditions and never the same condition order. Each participant heard only one of the lists (you can find the exact distribution of the lists on the OSF platform [https://osf.io/57ztj/]).

Each oddball condition consisted of 1000 trials in total (850 standards and 150 deviants; the ISI was jittered between 450 and 545 msec to achieve the same stimulus onset asynchrony for all items, which was 1147 msec)

resulting in a presentation time of approximately 20 min per condition and approximately 1 hr and 20 min in total. The order of the trials was fully randomized with at least two consecutive standards between deviants to avoid deviant stimuli developing their own memory trace. The initial 15 standards of each condition were excluded from all subsequent analyses, as their presentation served the sensory-memory trace formation (e.g., Näätänen et al., 2019), resulting in 3940 trials (985 trials × 4 conditions). Figure 5 depicts a schematic illustration of the oddball conditions.
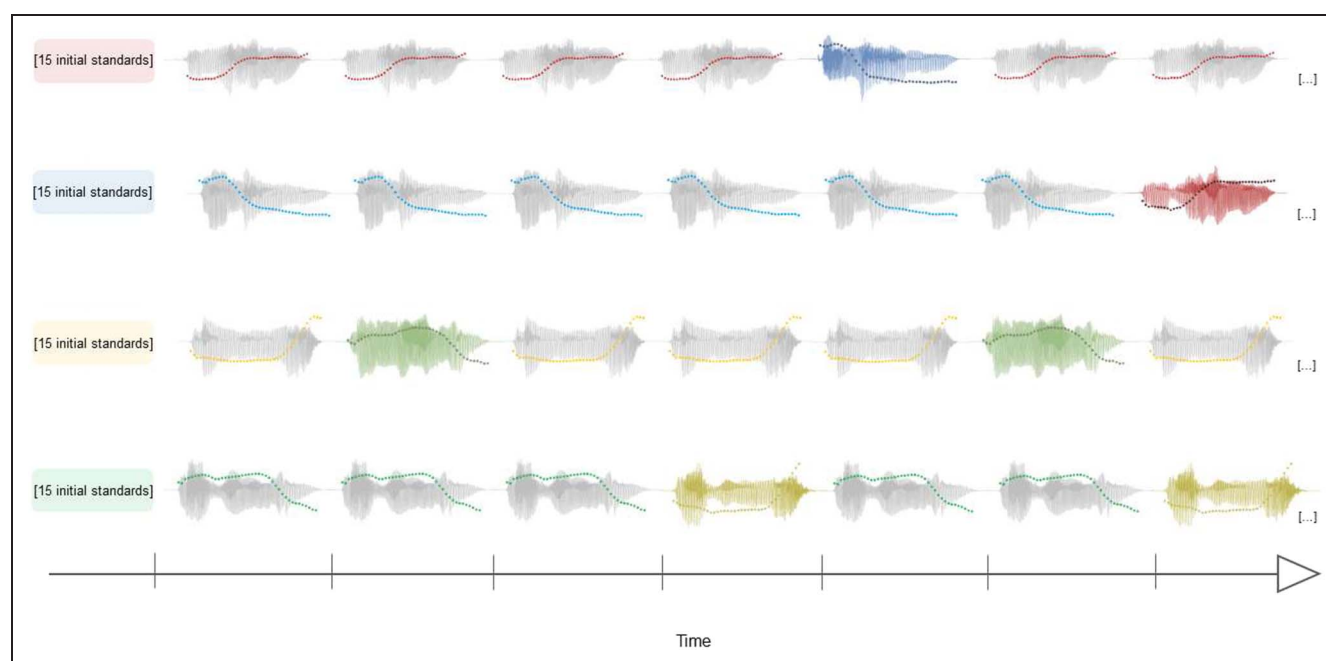
Figure 5. Schematic illustration of the oddball conditions. The four rows present the four different oddball conditions: From top to bottom, the first two rows illustrate the oddball conditions in which accentual rising (in red) and accentual falling (in blue) intonation alternate as standard/deviant within condition, whereas the last two rows depict the boundary rising (in yellow) and boundary falling (in green) oddball alternation within conditions. Each condition started with the presentation of 15 standard sounds serving the creation of the memory trace.

Data Preprocessing

The data were preprocessed using the MATLAB-based toolbox EEGLAB (Delorme & Makeig, 2004), developed at the Swartz Center for Computational Neuroscience. To reduce computational demands, the first step was to resample the data to 250 Hz. Afterward, the data were rereferenced to linked mastoids. Next, we performed an independent component analysis for artifact correction. For independent component analysis decomposition, the EEG was filtered with a 1-Hz high-pass filter to approach stationarity and a 45-Hz low-pass filter to remove line noise. Subsequently, artifact components (muscle and eye components above 80%; heart components above 90%) were automatically detected and removed from the raw EEG data. After artifact rejection, the raw EEG data were filtered with a 0.3-Hz high-pass and a 30-Hz low-pass filter, instead of baseline correction (cf. Maess, Schröger, & Widmann, 2016a, 2016b; Widmann, Schröger, & Maess, 2015; Wolff, Schlesewsky, Hirotani, & Bornkessel-Schlesewsky, 2008; Friederici, Wang, Herrmann, Maess, & Oertel, 2000). Thereafter, the data were epoched from −200 to 1000 msec post stimulus onset. For reproducibility, the preprocessing script can be found on the OSF platform (https://osf.io/57ztj/).

Postprocessing and Statistical Data Analysis

Postprocessing and statistical analyses were conducted in R, Version 4.1.2 (R Core Team, 2021). For data processing, we used the R package *tidyverse* 1.3.1 (Wickham et al., 2019), and for visualizations, we used the R package *ggplot2* 3.3.5 (Wickham, 2016). ERP amplitude was analyzed by fitting Bayesian hierarchical regression models using the *brms* 2.17.0 package (Bürkner et al., 2023). For reproducibility, data and scripts are available at https://osf.io/57ztj/ on the OSF platform.

Postprocessing

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

Our analysis focuses on two ERP effects that are claimed to index activation of pre-attentive and conscious attentional mechanisms, respectively: the MMN and the P3 responses (see Introduction). The MMN is a negative auditory ERP component that is traditionally obtained as a difference wave by subtracting the ERPs to standard from those to deviant stimuli (i.e., deviant ERPs – standard ERPs; Näätänen et al., 2019). However, this approach requires the use of the grand averaged signal, leading to a great loss of variance in the data. In our analysis,

we averaged ERP amplitude by time window for every participant, electrode site, and trial. This enables us to fit our models on single-trial data and at the same time model variance associated to each participant.

MMN is reported to typically peak between 100 and 250 msec after stimulus onset (e.g., Ducan et al., 2009). MMN is usually followed by a positive ERP component, the P3 response, around 300 msec or later after stimulus onset (P3 latency depends on the complexity of the processing, the more complex the processing, the longer the latency varying approximately from 250 to 1000 msec; e.g., Ducan et al., 2009). Nevertheless, there is considerable variability on the definition of the time windows that have been used to analyze these effects, as peak latency has been usually defined on the basis of difference waves. For example, there have been studies defining time windows with MMN peak latency at around 350 msec (e.g., Emmendorfer, Correia, Jansma, Kotz, & Bonte, 2020). For this reason, we analyzed ERP amplitude from 0 to 700 msec after stimulus onset in steps of 100 msec, resulting in seven time windows (i.e.: 0–100 msec, 100–200 msec, 200–300 msec, 300–400 msec, 400–500 msec, 500–600 msec, 600–700 msec). Furthermore, as the MMN is well documented to have a frontocentral topography, and the P3 a frontal distribution (e.g., Ducan et al., 2009), for our analyses, we defined a spatial region of interest[2] consisting of the AF3, AF4, F3, Fz, F4, FC1, FCz, FC2, and Cz electrode sites.

Inference Criteria

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

In the following section, we draw inferences using the posterior distributions of the parameters. For this, we report posterior estimates, the low and high boundaries of the 90% credible interval (CrI) of the estimate, and the posterior probability that the estimate falls on one side of zero (e.g., $P(\beta < 0) = 0.95$). When almost all of the posterior mass for an estimate lies on one side of zero, zero is
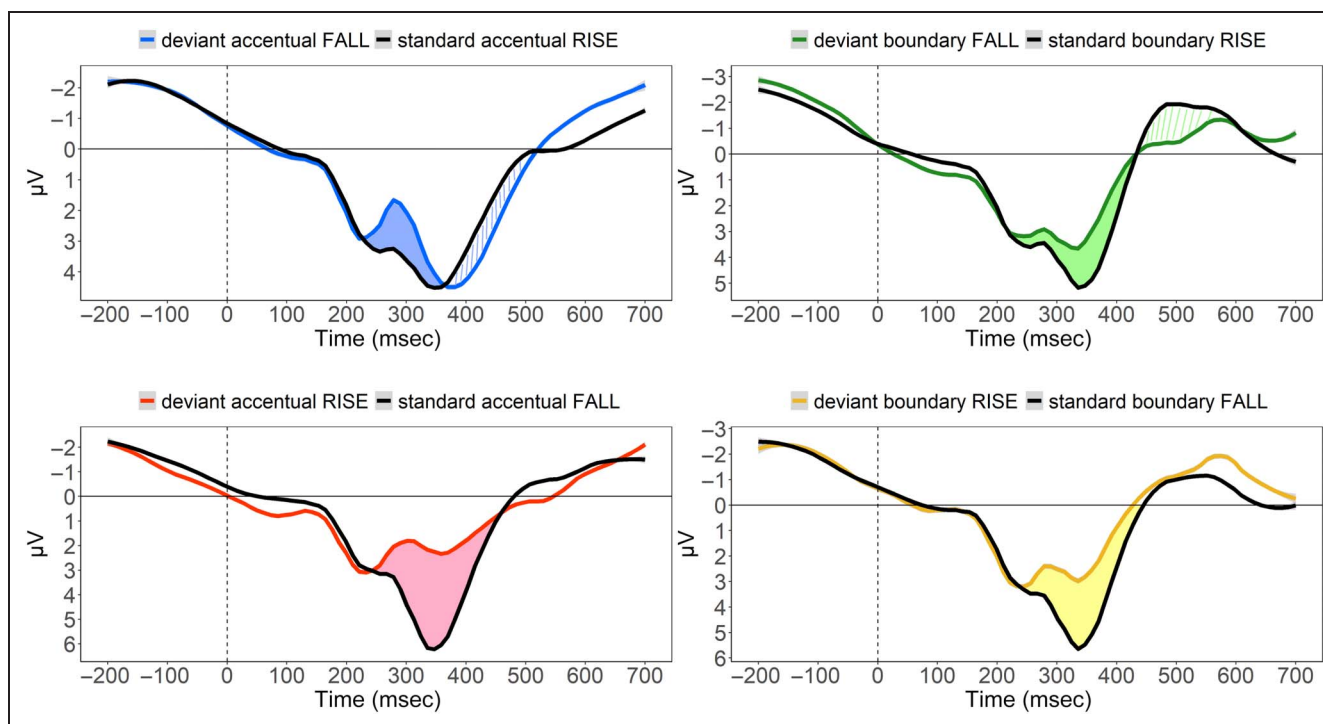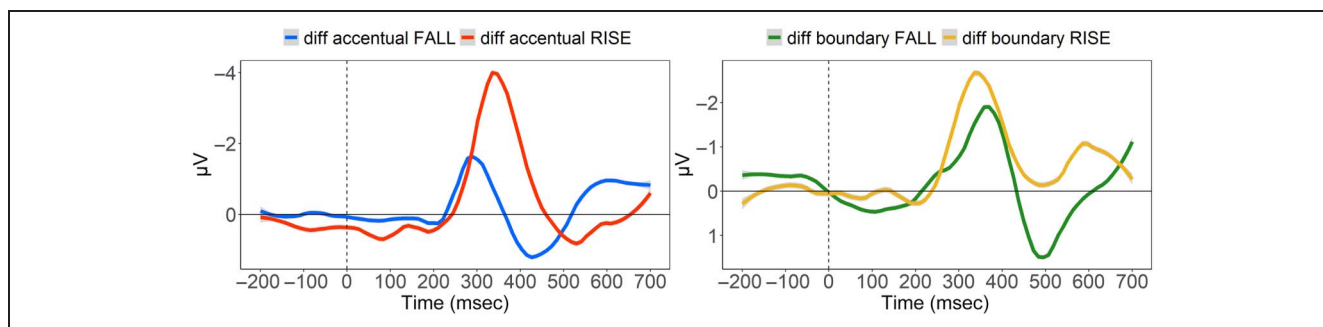
Figure 6. Grand average ERP waves (per oddball sequence) recorded to the onset of stimulus (illustrated by the vertical dashed line) over time (*x* axis) at the AF3, AF4, F3, Fz, F4, FC1, FCz, FC2, Cz electrode sites. Negative voltage is plotted upward. The left columns show ERPs to the two accentual oddball conditions in which accentual falls (in blue) and rises (in red) alternate as standard/deviant sounds. The right columns present ERPs with the two boundary oddball conditions in which boundary falls (in green) and rises (in yellow) alternate as standard/deviant sounds.

not included in the 90% CrI (by a reasonably clear margin), and the posterior probability *P* is close to one, we consider the effect reliable.

RESULTS

Figure 6 illustrates the grand averaged ERP waves per oddball condition time-locked to the onset of stimulus as depicted by the vertical dashed line (see also Figure 7 for grand averaged difference waves obtained by subtracting responses to standards from deviants per oddball condition). The left images show ERPs to the two accentual oddball conditions in which accentual falls (in blue) and

rises (in red) alternate as standard/deviant sounds. The right images present ERPs to the two boundary oddball conditions in which boundary falls (in green) and rises (in yellow) alternate as standard/deviant sounds. All images depict ERPs to standards in black, and EPRs to deviants in color. Visual inspection of the waves reveals that all contour types, when presented as deviants, evoked an MMN activity relative to their corresponding standard stimulation with an onset around 200 msec. For all deviants except the accentual fall (left top), the MMN activity (colored area between ERPs to standards and ERPs to deviants) appears to last for two successive time windows (200–400 msec), with the accentual rising deviant (left

Figure 7. Grand averaged difference waves time-locked to the onset of stimulus (illustrated by the vertical dashed line) and obtained in AF3, AF4, F3, Fz, F4, FC1, FCz, FC2, and Cz electrode sites. Difference waves were obtained by subtracting standard waves from deviant waves per oddball condition. The left side column depicts difference waves of accentual falling deviants (in blue) and accentual rising deviants (in red). The right side column shows difference waves of boundary falling deviants (in green) and boundary rising deviants (in yellow).

bottom) exhibiting the most pronounced effect. MMN to falling deviants (both accentual and boundary; top images) appears to be followed by an additional P3, at the 400- to 500-msec time window (shaded area between waves) for the accentual falls, and at a later time window (500–600 msec) for boundary falls.

In what follows, we first report results on modeling the difference between deviants and standards (i.e., standard sound vs. deviant sound) per oddball condition to detect whether MMN and P3 responses were elicited by deviant sounds relative to their standard stimulation. We refer to a brain activity as an MMN when we find a reliable negative difference between deviant and standard sounds during the 100–200 msec, 200–300 msec, and 300–400 msec time windows. We identify a brain response as P3 activity when we observe a reliable positive difference between deviant and standard stimulation in the 300–400 msec or at a later time windows. Subsequently, we summarize MMN and/or P3 effects (if any) in the presence of rises versus falls, regardless of their position (accentual/boundary) aiming to find whether the direction of the pitch movement affects the evoked brain response. Lastly, we sum up MMN and/or P3 effects in the light of the position of the rise and fall to find whether the phonological association of the tonal event affects the elicitation of MMN and/or P3 responses.

Accentual Contours

Posterior distributions of the estimated effects for the difference between standard and deviant sounds in the two accentual oddball conditions are shown in Figure 8. Blue color illustrates the estimated differences between accentual falling deviants and accentual rising standards (oddball condition1); red color depicts the estimated differences between accentual rising deviants and accentual falling standards (oddball condition2). Time windows are presented in ascending order.

Accentual Falling Deviant versus Accentual Rising Standard

Comparing the ERP amplitude between deviant accentual falling contours and standard accentual rising contours in the different time windows, the model revealed no reliable differences between 0–100 msec and 100–200 msec. Between 200 and 300 msec, although the 90% CrI includes zero on the margin, the model still very strongly favors the interpretation of a negative-going difference in amplitude, indicating an MMN activity. The model also estimated a reliable positive difference between 400 and 500 msec, indicating that MMN to falling deviants was followed by an additional P3, as well as another negative-going
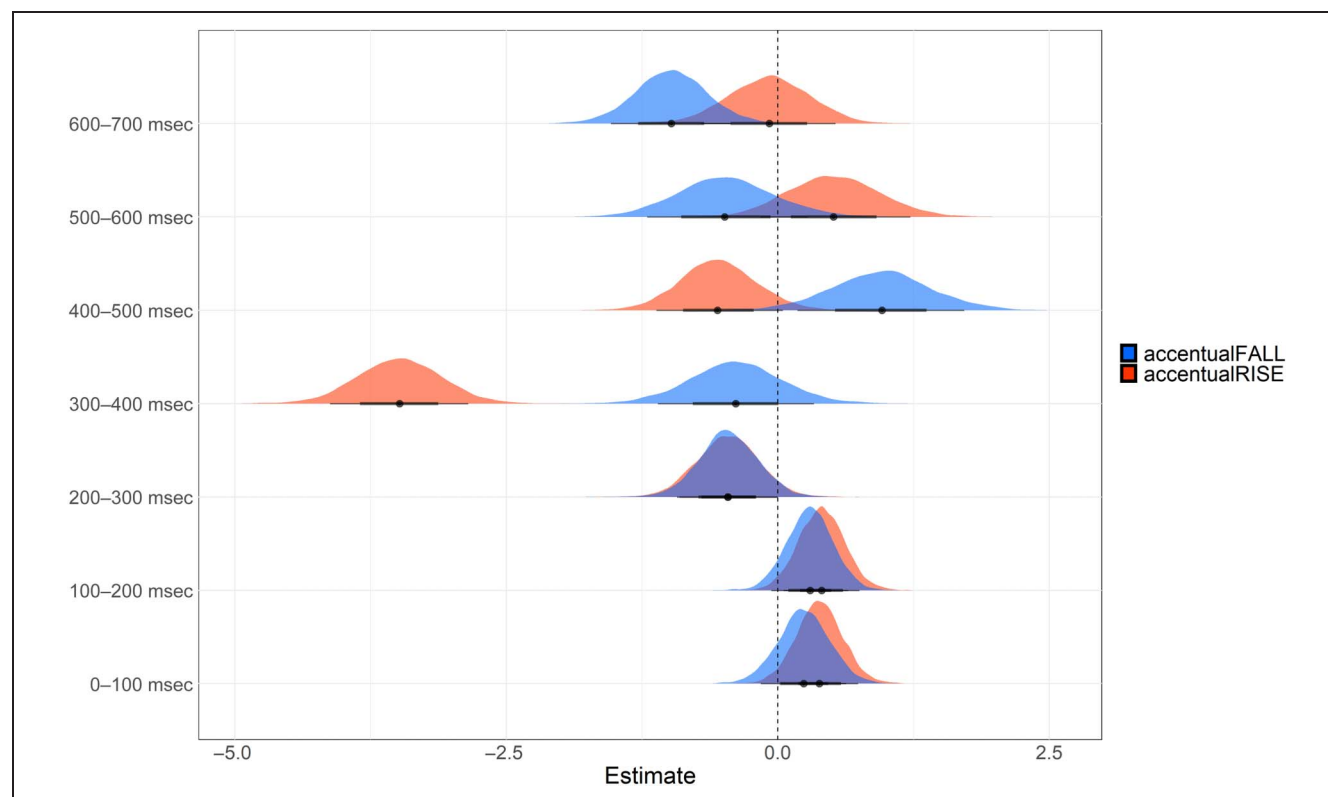
Figure 8. Posterior distributions of the estimated effects for the difference between standard and deviant sounds in the two accentual oddball conditions. Accentual falling deviants versus accentual rising standards are depicted in blue. Accentual rising deviants versus accentual falling standards are shown in red. Time windows are presented in ascending order. Error bars around the posterior means represent 66% (thick) and 90% CrI.

Table 2. Overview of Modeling Results for Accentual Falling Deviants versus Accentual Rising Standards per Time Window

Time Window	β	SE	Low CrI	High CrI	Evid.Ratio	Post.Prob
0–100 msec	0.24	0.24	−0.16	0.63	5.46	$P(\beta > 0) = 0.85$
100–200 msec	0.30	0.21	−0.06	0.65	10.89	$P(\beta > 0) = 0.92$
200–300 msec	−0.46	0.27	−0.90	0	19.78	$P(\beta < 0) = 0.95$
300–400 msec	−0.39	0.43	−1.1	0.33	4.61	$P(\beta < 0) = 0.82$
400–500 msec	0.95	0.46	0.18	1.72	44.2	$P(\beta > 0) = 0.98$
500–600 msec	−0.48	0.45	−1.2	0.28	6.04	$P(\beta < 0) = 0.82$
600–700 msec	−0.98	0.33	−1.54	−0.43	532.33	$P(\beta < 0) = 1$

difference between 600 and 700 msec. The model details are presented in Table 2.

Accentual Rising Deviant versus Accentual Falling Standard

For the contrasts between accentual rising deviants and accentual falling standards, the model estimated a reliable positive difference in amplitude in the 0- to 100-msec and 100- to 200-msec time windows. For the next two successive time windows, 200–300 msec and 300–400 msec, we found compelling evidence for a negative difference in amplitude, suggesting the presence of an MMN activity. For the remaining time windows, the model did not suggest any reliable amplitude difference. Table 3 shows the model details.

Overall, we find that both accentual rises and falls, when presented as deviant sounds, elicited an MMN activity starting in the 200- to 300-msec time window. For the accentual rising deviants, the MMN activity appears to have a longer duration than it does for the accentual falling deviants; as in this condition, a negative difference is present for two successive time windows (i.e., from 200 to 400 msec). For the accentual falling deviants, we find that its MMN activity is followed by an additional P3 response in the 400- to 500-msec time window, yet accentual rising deviants do not engender such a brain response.

Boundary Contours

Posterior distributions of the estimated effects for the differences between standard and deviant sounds in the two boundary oddball conditions are shown in Figure 9. The estimated differences between boundary falling deviants and boundary rising standards are illustrated in green (oddball condition3), whereas the estimated differences between boundary rising deviants and boundary falling standards are shown in yellow (oddball condition4). Time windows are presented in ascending order.
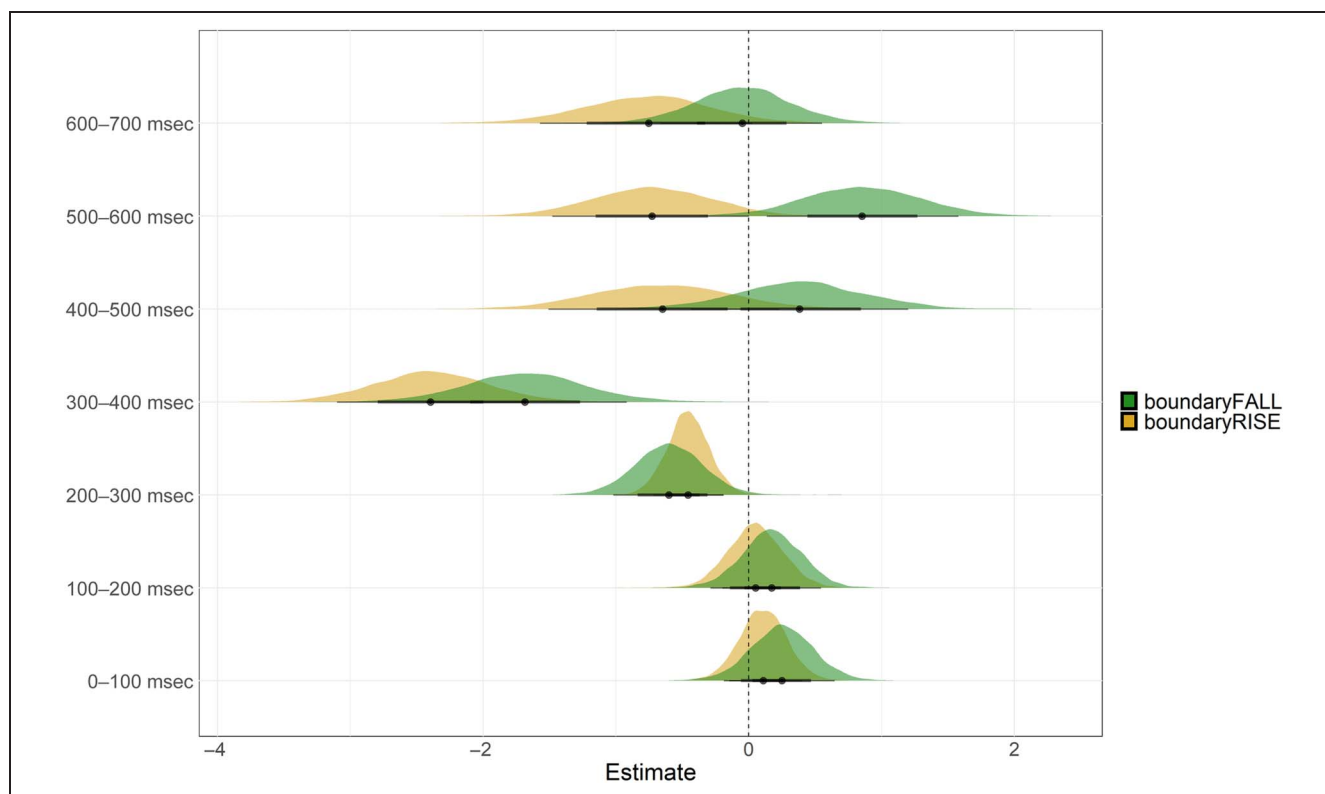
Boundary Falling Deviant versus Boundary Rising Standard

Comparing the ERP amplitude between boundary falling deviant and boundary rising standard contours, there were no reliable differences in the 0- to 100-msec and 100- to 200-msec time windows, but we found compelling evidence for a negative difference in amplitude between 200–300 msec and 300–400 msec, indexing the elicitation of an MMN activity. Between 400–500 msec, the model estimated a positive difference, yet this difference was not reliable. In the 500- to 600-msec time window, the model provided compelling evidence for another positive-going difference, suggesting the presence of a P3 response. Between 600 and 700 msec, no reliable differences were reported. The model details are presented in Table 4.

Table 3. Overview of Modeling Results for Accentual Rising Deviants versus Accentual Falling Standards per Time Window

Time Window	β	SE	Low CrI	High CrI	Evid.Ratio	Post.Prob
0–100 msec	0.38	0.22	0.03	0.74	25.58	$P(\beta > 0) = 0.96$
100–200 msec	0.40	0.21	0.06	0.75	34.56	$P(\beta > 0) = 0.97$
200–300 msec	−0.46	0.29	−0.93	0	18.51	$P(\beta < 0) = 0.95$
300–400 msec	−3.49	0.39	−4.13	−2.83	Inf	$P(\beta < 0) = 1$
400–500 msec	−0.55	0.36	−1.12	0.05	14.78	$P(\beta < 0) = 0.94$
500–600 msec	0.52	0.42	−0.16	1.22	8.22	$P(\beta > 0) = 0.89$
600–700 msec	−0.08	0.38	−0.71	−0.53	1.42	$P(\beta < 0) = 0.59$

Figure 9. Posterior distributions of the estimated effects for the differences between standard and deviant sounds in the two boundary oddball conditions. Boundary falling deviants versus boundary rising standards are depicted in green. Boundary rising deviants versus boundary falling standards are shown in yellow. Time windows are presented in ascending order. Error bars around the posterior means represent 66% (thick) and 90% CrI.

Boundary Rising Deviant versus Boundary Falling Standard

For the contrast between boundary rising deviants and boundary falling standards, we found no evidence for a difference in the 0- to 100-msec and 100- to 200-msec time windows, yet between 200–300 msec and 300–400 msec, the model provides compelling evidence for a negative difference in amplitude, indicating an MMN activity. Similarly, for the last three time windows (400–500 msec, 500–600 msec,

and 600–700 msec), the model estimates a negative-going difference, but this difference is not reliable between 400 and 500 msec, whereas between 500–600 msec and 600–700 msec, although the 90% CrI includes some zero on the margin, it could still favor the interpretation of a (marginal) difference. Table 5 presents model details.

In summary, for boundary contours, we find that both rises and falls evoke an MMN activity relative to their corresponding standard stimulation, lasting from 200 to 400 msec. Furthermore, brain responses to boundary rises

Table 4. Overview of Modeling Results for Boundary Falling Deviants versus Boundary Rising Standards per Time Window

Time Window	β	SE	Low CrI	High CrI	Evid.Ratio	Post.Prob
0–100 msec	0.25	0.24	−0.15	0.65	5.93	$P(\beta > 0) = 0.86$
100–200 msec	0.17	0.23	−0.20	0.54	3.68	$P(\beta > 0) = 0.79$
200–300 msec	−0.60	0.25	−1.02	−0.19	107.11	$P(\beta < 0) = 0.99$
300–400 msec	−1.68	0.46	−2.43	−0.92	1599	$P(\beta < 0) = 1$
400–500 msec	0.39	0.49	−0.43	1.20	3.82	$P(\beta > 0) = 0.79$
500–600 msec	0.85	0.44	0.14	1.58	34.4	$P(\beta > 0) = 0.97$
600–700 msec	−0.05	0.37	−0.66	0.55	1.24	$P(\beta < 0) = 0.55$

Table 5. Overview of Modeling Results for Boundary Rising Deviants versus Boundary Falling Standards per Time Window

Time Window	β	SE	Low CrI	High CrI	Evid.Ratio	Post.Prob
0–100 msec	0.11	0.18	−0.19	0.40	2.75	$P(β > 0) = 0.73$
100–200 msec	0.05	0.21	−0.29	0.39	1.49	$P(β > 0) = 0.60$
200–300 msec	−0.46	0.16	−0.72	−0.19	379.95	$P(β < 0) = 1$
300–400 msec	−2.39	0.43	−3.1	−1.68	Inf	$P(β < 0) = 1$
400–500 msec	−0.65	0.53	−1.51	0.23	8.36	$P(β < 0) = 0.89$
500–600 msec	−0.73	0.45	−1.48	0	18.75	$P(β < 0) = 0.95$
600–700 msec	−0.77	0.48	−1.57	0.03	17.43	$P(β < 0) = 0.95$

and falls differ in that the MMN to boundary falls is followed by a P3 response, yet there is no evidence for boundary rises eliciting a subsequent positivity.

Interim Summary

Figure 10 depicts all posterior distributions of the estimated effects for the differences across all four oddball conditions. Table 6 presents an overview of the estimated differences between rising and falling contours as well as between accentual and boundary contours. These differences point toward a distinct relevance of contour direction (i.e., rise or fall) and linguistic context for the attentional mechanisms.

Comparing accentual rising to falling deviants, we find that both contours elicit an MMN activity with an onset in the 200- to 300-msec time window. There is no quantitative difference in the beginning of the MMN elicitation between accentual rises and falls, yet MMN to accentual rises is prolonged over the subsequent time window (300–400 msec), exhibiting thus a longer-lasting effect compared with the accentual falls. In addition, accentual falls after MMN elicitation engender a subsequent P3 in the 400- to 500-msec time window, a response that is not evoked by accentual rises. Turning to the rising versus

falling comparison in the boundary conditions, the findings are quite similar. Both boundary rises and falls evoke an MMN activity lasting for two successive time windows (200–400 msec), with a tendency for the boundary rises to show a more negative effect in the 300–400 msec. The evidence from the model is not reliable enough to claim with confidence that the MMN to rises is more pronounced than MMN to falls. However, MMN to boundary falls, similarly to MMN to accentual falls, is followed by an additional P3. By contrast, there is no evidence for boundary rises eliciting a subsequent positivity.

Considering the position of the rise and fall, MMN to falling contours may be of a similar magnitude in the starting time window (200–300 msec), regardless of it being accentual or boundary, yet MMN to boundary falls is prolonged over the subsequent time window (300–400 msec) as opposed to the activity evoked by the accentual falls. Furthermore, MMN to falling deviants is followed by an additional P3, at the 400- to 500-msec time window for the accentual falls, and at a later time window (500–600 msec) for boundary falls. Turning to rising contours, the onset of the MMN activity does not differ as a function of the phonological status of the rise (accentual vs. boundary), but the accentual rises exhibit the most pronounced effect in the 300- to
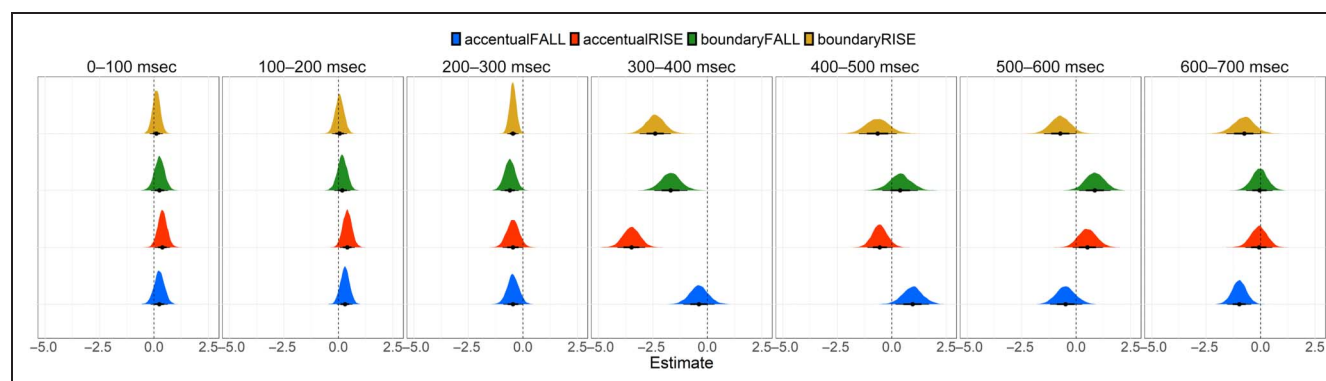
Figure 10. Posterior distributions of the estimated effects for the differences between standard and deviant sounds across oddball conditions. From top to bottom, boundary rising deviants versus boundary falling standards are shown in yellow, boundary falling deviants versus boundary rising standards are depicted in green, accentual rising deviants versus accentual falling standards are illustrated in red, and accentual falling deviants versus accentual rising standards are presented in blue. Time windows of interest from left to right are the 200–300 msec, 300–400 msec (MMN time windows), and the 400–500 msec, 500–600 msec (P3 time windows). Error bars around the posterior means represent 66% (thick) and 90% CrI.

Table 6. Overview of Estimated Differences between Rising and Falling Contours as well as Accentual and Boundary Contours in the Time Windows of Interest

Time Window	β	SE	Low CrI	High CrI	Evid.Ratio	Post.Prob
Rises vs. Falls—Accentual						
MMN (200–300 msec)	0	0.39	−0.65	0.64	1.02	$P(\beta < 0) = 0.5$
MMN (300–400 msec)	−3.1	0.59	−4.07	−2.15	Inf	$P(\beta < 0) = 1$
P3 (400–500 msec)	1.5	0.59	0.54	2.47	132.33	$P(\beta > 0) = 0.99$
Rises vs. Falls—Boundary						
MMN (200–300 msec)	0.14	0.30	−0.35	0.64	0.46	$P(\beta < 0) = 0.31$
MMN (300–400 msec)	−0.71	0.62	−1.74	−0.28	7.11	$P(\beta < 0) = 0.88$
P3 (500–600 msec)	1.58	0.63	0.56	2.61	144.45	$P(\beta > 0) = 0.99$
Accent vs. Boundary—Rises						
MMN (200–300 msec)	0	0.33	−0.56	0.53	1.01	$P(\beta < 0) = 0.5$
MMN (300–400 msec)	−1.1	0.58	−2.05	−0.15	36.21	$P(\beta < 0) = 0.97$
Boundary vs. Accent—Falls						
MMN (200–300 msec)	−0.14	0.37	0.37	0.46	1.89	$P(\beta < 0) = 0.65$
MMN (300–400 msec)	−1.29	0.63	−2.36	−0.25	53.79	$P(\beta < 0) = 0.98$

400-msec time window. Finally, we find no evidence for a P3 brain response in the presence of either accentual or boundary rises.

We return to these findings in the Discussion where we argue that the brain responses that we observe for the tested contours could potentially indicate different neurocognitive processes of speech rises and falls, both because of speech sound complexity and linguistic or context-specific interpretation.

DISCUSSION

Unlike previous research on neurophysiological correlates of unexpected changes in a stream of repetitive stimulation that focused primarily on nonlinguistic stimuli, our work explores the neural responses to rising and falling pitch attributable to accents and boundary tones of sequentially presented lexical items in German. The main aim of this study was twofold. Using EEG data in an oddball paradigm, we investigated the role of linguistic rises in attention orienting, as well as how far the phonological status of the rise plays a role.

It is important to note that we discuss only MMN- and P3-related effects, as these are the brain responses that relate to our research questions. First, we put our findings in the broader context of rising versus falling intonation in language (irrespectively of the prosodic structure), and their relevance for attention orienting (Hypothesis 1). Subsequently, we discuss whether the phonological status

of the rise and the fall (accent vs. boundary) modulates attentional resources (Hypothesis 2).

The Processing of Rises and Falls and Their Contribution to Attention Orienting

In our results, we find evidence for distinct neurocognitive processes of intonational rises and falls in the context of an oddball list. This is reflected in the different brain responses that we observe during the online processing of our tested pitch contours: Rising pitch contours engendered MMN activity, whereas falling pitch contours evoked an MMN-P3 complex. Whereas MMN indexes automatic/preconscious processes related to involuntary attention switch toward the deviant pitch contour, the MMN-P3 complex indicates that processes at the preattentive level subsequently activate processes at the conscious level, bringing the perception of the deviant contour into awareness and voluntary attention.

These processing patterns might reflect the presence of different mechanisms or routes for signal-driven (bottom–up) and context-driven (top–down) processes. The former have been the target of previous research utilizing the oddball paradigm, whereas the presentation of lexical items in the present study has the potential to give rise to contextual expectations for a particular list intonation. Our data indicate that in a linguistic context, the acoustic signal is not the only source for attention orienting, as would be expected in a "meaningless" context of pure sine waves.

Pitch like any other acoustic property is processed as the sensory input unfolds, while at the same time, expectations for the forthcoming input are formed incrementally (e.g., Röhr et al., 2021). In a linguistic context, the generation of expectations can be based both on the sensory information of the input (signal-driven) as well as on the contextual meaning (context-driven; Röhr et al., 2021). This last point deserves special attention, and it is thus worth elaborating on how our results elucidate this point.

In two of our oddball conditions, falling pitch contours were presented as deviant sounds, whereas their corresponding standard stimulation consisting of repetitive rises. In such sequences, the listener can potentially build predictions derived from two different sources. First, the auditory processing system is able to predict prospective sounds by detecting regularities in the sensory input (see section Attention Orienting toward Unexpected Sound Events). Thus, the listener, after being exposed to a repetitive sequence of stimuli with rising pitch, predicts and anticipates that the next auditory event will again be a rising contour. Predictions can also arise from the linguistic interpretation of the context. The oddball paradigm resembles a list context because it presents (auditory) events repeatedly and sequentially. On top of the sequential presentation of our stimuli, we used naturalistic pitch contours realized on real words, which potentially simulated an even more natural list context (these words all refer to food items that could conceivably be used in lists such as shopping lists). The list intonation in German, as mentioned above, typically involves rising pitch on nonfinal and penultimate items (denoting continuity) followed by a fall on the final item (indicating finality; Peters, 2018; Baumann & Trouvain, 2001). Thus, the repetitive rising stimulation, as a natural and appropriate contour on nonfinal list items, denotes that the list is not over yet. The repetitive/standard rises, therefore, might have elicited additional predictions driven by the contextually created meaning. Such an expectation is the anticipation that the list will at some point in time end. Therefore, the listener, first, anticipates a rising contour on the basis of the sensory information that is available in the repetitive auditory stimulation and, second, given the available contextual meaning, expects that at some point the list will be over, anticipating thus a falling contour. Recall now that when the deviant fall was presented, an MMN-P3 complex was elicited, indicating that first, the violation of the anticipated rising contour activated a pre-attentive response of an involuntary attention switch to the unexpected falling contour. Subsequently, a conscious or voluntary attention orientation toward this falling contour was observed, potentially induced by the validation of the context-driven expectation, that is, the anticipation that the list would at some point come to an end.

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

To better understand how these distinct processes between the rising/falling deviants, as reflected in the brain responses we observe, link to attention orienting and prominence cueing, we need to consider our findings jointly. The starting observation is that all pitch contours when presented as deviants evoked an MMN activity, indexing orientation toward the violation in the acoustic signal. In particular, rising pitch deviants, either on a stressed syllable or at the boundary of a word engendered the most pronounced MMNs, showing that pitch rises attract the most involuntary attention. Falling pitch deviants elicited less pronounced MMNs followed by a pronounced P3, indicating that pitch falls also attract some involuntary attention which ultimately leads to conscious attention orientation. This suggests that rising pitch, as an acoustically prominent cue, causes an auditory looming effect at the pre-attentive stage, whereas falling pitch appears to be interpreted as linguistically prominent information within the list context; thus, its processing is affected by linguistic meaning, also activating conscious attentional mechanisms. Our results appear to differ from the Hsu and colleagues (2015) study in that they found a P3 response to rising pitch changes at a speaker's normal pitch level (i.e., not elevated).

In what follows, we suggest that our findings do not differ greatly from those of Hsu and colleagues. Remember, our stimuli were different from those used in Hsu and colleagues because we used real words as opposed to a simple /ɑ/ (at normal and resynthesized elevated pitch levels) and sine waves. An additional difference is that the rise or fall in the Hsu and colleagues stimuli involved a change in pitch from one stimulus to the other, thus across stimuli; whereas in our study, the rise or fall appears within stimuli. Hsu and colleagues suggest that sudden pitch rises in speech demand more attentional resources than sudden falls, and their presence in speech (in comparison to simple sine waves) activates additional conscious processing mechanisms. In our case, where the speech signal is the only signal that listeners encounter, we conjecture that it is the linguistic/contextual meaning

that draws voluntary attention rather than speech per se. Put differently, one could argue that in both Hsu and colleagues and in our study, P3 appears to be elicited by some kind of available "meaning" and not so much by the direction of the pitch. In Hsu and colleagues, "meaning" emerges from the speech in comparison to the meaningless sine waves. It is the sudden rising and not the sudden falling pitch that evokes the P3 because rising pitch is acoustically more prominent, but if it was only the direction of pitch and not the additional information of "meaning," then pitch rises in sine waves would have elicited a P3 as well. In our study, "meaning" arises from the context. The rising deviant condition, although acoustically prominent, happens to be contextually meaningless, like the sine waves in the Hsu and colleagues study, whereas the falling deviant condition, although acoustically less prominent, transmits the linguistic meaning of the list context. Such a list context may not be as available to speakers of Mandarin, where local f0 changes are affected by the lexical tone.

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

Overall, our results show that the processing of rising and falling pitch contours produce distinct brain responses that are claimed to be related to two core attentional processes: the automatic involuntary attention switch at the pre-attentive stage (reflected by the MMN generation) and the voluntary attention orientation at the conscious stage (reflected in the P3 signature). Pre-attentively, signal-based cues appear to be fundamentally important. At this processing stage, pitch rises, as the most acoustically prominent events, attract the most attention (for similar findings, see Li & Chen, 2018; Ren, Yang, &

Li, 2009, among others).[3] This highlights the pivotal role of pitch rises not only in cognition, as previous MMN studies have shown, but also in language, as it appears that there is something intrinsic in their acoustic signal, regardless of whether they are a pure sine wave, a speech sound in isolation or speech in context. This is potentially so because the rising acoustic properties are so prominent that they are able to warn and prepare listeners' nervous system about an important event happening in the environment, activating basic attentional resources that in turn elicit automated or appropriate adaptive responses (these adaptive responses have been described by Sokolov, 1963 as reflexes). In spoken communication, this could entail orienting a listener's attention toward the most important part of the uttered message, which is crucial for effective interpretation and speech planning, as drawing a listener's attention to an upcoming turn. Nevertheless, our findings manifest that conscious, voluntary attention is modulated by the meaning that intonation encodes in a given context and not by the pitch direction itself (see also Liu et al., 2016, for similar attentional processes modulated by context-driven cues). In our study, the conscious processing stage is activated by pitch falls. Although pitch falls are the least prominent events in the prosodic prominence hierarchy, it is evident that, in our case, language experience and context-driven expectations overwrite the signal-induced properties (see Bishop, 2013). Put differently, the signal-based properties, cueing acoustic prominence, can be replaced by expectations that emerged from context, making pitch falls highly relevant in the list context.

The Role of the Phonological Status of the Rise (and the Fall) in Attention Orienting

Another question we have put forward in this study concerns the phonological status of the rises and falls, that is, their position in the prosodic structure and its contribution to the attentional processes we have observed. Crucially, we only observed differences in magnitude and latency of the MMN activity when comparing accentual to boundary contours (see Figure 10). Specifically, when comparing accentual to boundary rises, we find that accentual rises exhibit the most pronounced MMN. In turn, when comparing accentual to boundary falls, we observe that boundary falls show a prolonged MMN latency as opposed to the very short activity evoked by accentual falls. We argue that during online processing, intonation contours, as a complex part of the speech signal, are processed holistically, meaning that attention is not oriented necessarily toward a specific point in the f0 trajectory. Rather, it is the pitch contour that modulates attention orienting (drawing attention to a higher level of linguistic representation such as a—putative—phrase).

To illustrate this point, consider the investigated rising/falling contours: Both accentual and boundary contours were realized on sequentially presented trisyllabic lexical items (like *Banane* "banana"). Hence, the domain

of the realization of the pitch contour in this study is the word; thus, each word is an intonational phrase on its own. Thus, a complete pitch contour on every word consists of both a pitch accent and a boundary tone. Specifically, the pitch configuration, which in this study we call *accentual rise*, is followed by a high boundary at the end of the word, whereas the pitch configuration, which we call *accentual fall*, is followed by a low boundary at the end of the word. Likewise, our *rising boundary* is preceded by a low accent, whereas our *falling boundary* is preceded by a high accent. Hence, considering the entire pitch configuration as our word stimuli unfold, we can better understand whether a specific part of the contour is relevant for attracting (more) attention or not. Our results indicate that the structural position of the pitch event (accent vs. boundary) has a secondary role in attentional processes. Specifically, we find that (i) a rising pitch contour is globally more successful in attracting attention than a falling one, regardless of the position of the rise or the fall, (ii) whereas secondarily, within rising and falling contours, if there is a rise on the stressed syllable (rather than on the final unstressed syllable), this pitch configuration leads to more attention.

The finding that accentual rises induce a greater MMN effect, attracting more attention than boundary rises, is not surprising. It has already been claimed in previous work on prosodic prominence that accentual rising contours (and especially steep rising pitch accents) are the most prominent contours in the prosodic prominence hierarchy and that they demand more attentional resources than falling pitch accents (Röhr et al., 2021; Baumann & Röhr, 2015). Let us now consider our finding (accentual rises attracting more attention than boundary rises) in light of the entire contour and the periodic energy that characterizes these rising contours (for results on periodic energy, see section 2.2.1). In our accentual rising contours, the pitch starts low, already rises quite steeply during the stressed syllable, and remains high toward the last syllable and, thus, at the end of the word (an appropriate German Tones and Break Indices analysis [Grice, Baumann, & Benzmüller, 2005] would be: L + H* H-%). The rising pitch movement is on the lexically stressed syllable, which has high periodic energy, making the pitch strongly transmitted on that syllable. See Figure 1 for mean and individual intonation contours, and Figures 2 and 3 for measures related to periodic energy. Now, in boundary rising contours, the pitch starts at a mid-level, remains at this level or falls slightly during the stressed syllable, and rises toward the end of the word. In GToBI, this is L* L-H%. The rising part of the contour is restricted to the final unstressed syllable, which has considerably lower periodic energy than the stressed syllable. Thus, the f0 in the boundary rise condition is more weakly transmitted than in the accentual rise condition. Therefore, although both rising contours attract more attention than the falling ones, within the rising category, the accentual rising contours are produced with more energy

and thus attract more attention than the boundary rising contours.

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

One could argue that this is evidence for the prominence value of the accent and thus the structural importance of the stressed syllable. However, remember that when we compare boundary rising to boundary falling contours, we find that the boundary rises attract more attention than falls, although the stressed syllable in the former contour bears a low pitch accent, as opposed to the boundary fall that is preceded by a high pitch accent. It appears therefore that what happens (in terms of f0 movement) at discrete prosodic positions (head/edge) is not sufficient to orient attention on its own. In contrast, it is the holistic processing of the entire contour that guides attention. Prosodic positions appear to have a supplementary/secondary role in the modulation of attentional resources. This becomes evident when we arrange the investigated pitch contours according to their elicited MMN effects, assuming a decrease in MMN effect, and thus attention attraction, from left to right:

RISE[accentual rises (L + H* H-%) > boundary rises (L* L-H%)].
> **FALL**[(boundary falls (H* L-%) > accentual falls (H + L* L-%)].

First and foremost, a rising pitch configuration attracts globally more attention than a falling one. Second, and within rising categories, when the rise occurs in different structural positions, it attracts more attention when it coincides with the head/stressed syllable (in the accentual rising condition) compared with when it occurs at the boundary (in the boundary rising condition). Now, when the rise occurs at the same structural position, it attracts more attention if it is a steep rise (L + H* in the accentual rising condition) as opposed to a shallow or just a

high pitch (H* in the boundary falling condition). Finally, within falling categories, a shallow rise preceding a fall on the accented syllable (in the boundary falling condition) is better in attracting attention than a simple fall (in the accentual falling condition). These results are in line with what has been previously found about prominence marking in German (e.g., Baumann & Winter, 2018). Crucially, here, we show that these subtle prosodic differences are reflexed in the pre-attentive MMN response.

Conclusion

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

I. the involuntary attention switch, a mechanism at the pre-attentive processing stage, reflected in the MMN elicitation,
II. the voluntary attention orientation, a mechanism at the conscious processing stage, reflected in the P3 signature.

Some of the cues in our stimuli have primary relevance for the two attentional mechanisms, and some have only secondary relevance. We find that the pitch contour itself (signal-driven cue) and the appropriateness of the contour in the linguistic context (context-driven cue) are decisive for the two attentional mechanisms, involuntary and voluntary, respectively, whereas the phonological status of the pitch event (head/edge) plays only a supplementary role.

In its most concise form, our main finding is that, in spoken language, rising intonation takes on a special role in attracting involuntary attention, whereas contextual meaning is essential for voluntary attention orienting. Rising pitch evokes the largest MMN, indicating that it leads to a greater involuntary and automated attention switch compared with falling pitch. This holds, regardless of the phonological association of the rise in the prosodic structure (head vs. edge). It appears thus that there is a biological basis for cross-linguistic use of rises for attracting attention toward informative parts of the message. Although they will be grammaticalized differently across languages, that is encoded as pitch accents or edge tones. Furthermore, the appropriateness of the intonational pattern in the context and, specifically in our case, in the list

context, is decisive for voluntary attention orienting. In our study, falling pitch, although acoustically less prominent than rising pitch, engenders an additional P3, indicating that the contextual meaning prevails over or even "cancels out" the signal properties (for a discussion on the interplay between sensory input and top–down activities in language processing, see Bornkessel-Schlesewsky et al., 2022).

To conclude, the neural architecture of language perception is complex and dynamic (for discussions, see Assaneo et al., 2019; Hickok & Poeppel, 2007; Gandour et al., 2004, among others), involving two fundamentally different neural mechanisms, a signal-based mechanism and a meaning-dependent one, expressed at distinct processing stages. On the one hand, our results show that the intrinsic properties of the sensory input, that is, signal-driven cues (Assaneo et al., 2019, refers to it as the *intrinsic auditory mechanism*), are essential for speech perception at the early pre-attentive processing stage, with rises taking priority over falls precisely because of their acoustic prominence. The fundamental role of signal properties, and thus the special role of rises, at the pre-attentive processing stage (feeding thus the signal-based mechanism) is also shown in previous MMN studies to sound changes (for a review of studies, see Näätänen et al., 2019) but also in studies investigating the neural processing of linguistically meaningful variations both at lexical and postlexical levels (e.g., Li & Chen, 2018; Ren et al., 2009, among others). On the other hand, our findings suggest that at a later, conscious, processing stage, evolving after the pre-attentive one, the linguistic functions of the stimuli (called *top–down/externally driven mechanism* by Assaneo et al., 2019) modulate speech perception (feeding thus the meaning-dependent mechanism). We find that the construction of meaning attracts voluntary attention toward meaningful aspects, here reflected by the use of words and the minimal context of list intonation. This is in line with findings from previous studies showing that top–down activities are decisive for the activation of later processing stages (e.g., Assaneo et al., 2019; Liu et al., 2016; Ren et al., 2009).

Overall, the role of rises is fundamentally important, not only for cognition but also for (successful) language communication, because of their intrinsic acoustic properties that activate involuntary attentional resources, pre-attentively, regardless of them characterizing a simple sine wave, a speech sound in isolation, or meaningful speech. However, in language, the acoustic signal is not the only source of information; hence, their importance is mitigated by the contextual meaning, which appears to activate voluntary attention, as it is required for modulating/updating conscious processing stages and mental representations during language comprehension.

APPENDIX—PRIOR SPECIFICATIONS

We used weakly informative priors for all our parameters, as they allow for a wide range of effect sizes but control for unreasonable large effects:

- For the intercept, we assumed a normal distribution with mean 0 and standard deviation 3.
- For the fixed effect, we assumed a normal distribution with mean 0 and standard deviation 2.
- For the random effects standard deviations, we assumed a truncated (takes only positive values) normal distribution with mean 0 and standard deviation 2.
- For the residual noise (sigma), we assumed a truncated normal distribution with mean 0 and standard deviation 30.
- For the variance–covariance matrix of the random effects, we assumed a Lewandowski, Kurowicka, and Joe (2009), prior with parameter value 2.

Corresponding author: Maria Lialiou, Department of German Language and Literature I, Linguistics, University of Cologne, Luxemburger Str. 299, Cologne, 50939, Germany, or via e-mail: mlialiou@uni-koeln.de.

Data Availability Statement

Data and scripts for all analyses are available online on the OSF platform under this link: https://osf.io/57ztj/.

Author Contribution

Maria Lialiou: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Validation; Visualization; Writing—Original draft; Writing—Review & editing. Martine Grice: Conceptualization; Funding acquisition; Investigation; Methodology; Project administration; Supervision; Writing—Review & editing. Christine T. Röhr: Conceptualization; Methodology; Writing—Review & editing. Petra B. Schumacher: Conceptualization; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Writing—Original draft; Writing—Review & editing.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

Notes

1. The frequency distributions of the words were extracted from the Leipzig Deutscher Wortschatz corpus (Quasthoff & Richter, 2005).
2. At first, we intended to run our models using all registered electrodes including sagittality and laterality as predictors in the models. However, except for the theoretical reasons we report here, there were also practical reasons that led us to the decision of defining one spatial ROI. That is, the models including all scalp electrode sites were highly computationally demanding, that is, they were computationally intractable.
3. The Li and Chen study investigated T2 (rising) and T4 (falling) tones, similar to our intonational rising and falling contours. Using the MMN paradigm, they compared the timing of the cues by contrasting T2/T4 with T3 as reference tone and found that MMN is sensitive to the timing of the acoustic cues (cue of divergence point). Of particular interest is that the MMN time window they found for the contours with the early divergence cue is the same time window that we observed in our study. Our contours also have an early divergence point (i.e., divergence from the beginning of the word in our stimuli), pointing at a cross-linguistic similarity. Similar findings are reported by Liu and colleagues (2016), which is again on the processing of tone and intonation in Mandarin Chinese. Crucially, they observe that P3 is modulated by the context (question vs. statement) for the falling contour T4.

REFERENCES

Alain, C., Woods, D. L., & Ogawa, K. H. (1994). Brain indices of automatic pattern processing. *NeuroReport*, 6, 140–144. https://doi.org/10.1097/00001756-199412300-00036, PubMed: 7703403

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

Albert, A., cangemi, f., Grice, M., & Ellison, T. M. (2023). ProPer: PROsodic analysis with PERiodic energy. https://doi.org/10.17605/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. *Frontiers in Integrative Neuroscience*, *13*, 28. https://doi.org/10.3389/fnint.2019.00028, PubMed: 31379527

Bach, D. R., Schächinger, H., Neuhoff, J. G., Esposito, F., di Salle, F., Lehmann, C., et al. (2008). Rising sound intensity: An intrinsic warning cue activating the amygdala. *Cerebral Cortex*, *18*, 145–150. https://doi.org/10.1093/cercor/bhm040, PubMed: 17490992

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001, PubMed: 24403724

Baumann, S., & Röhr, C. T. (2015). The perceptual prominence of pitch accent types in German. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)* (Vol. 298, pp. 1–5). Glasgow, UK: The University of Glasgow.

Baumann, S., & Schumacher, P. B. (2012). (De-)accentuation and the process of information status: Evidence from event-related brain potentials. *Language and Speech*, *55*, 361–381. https://doi.org/10.1177/0023830911422184, PubMed: 23094319

Baumann, S., & Trouvain, J. (2001). On the prosody of German telephone numbers. In *7th European conference on speech communication and technology (Eurospeech 2001)* (pp. 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

Bishop, J. (2013). Information structural expectations in the perception of prosodic prominence. In G. Elordieta & P. Prieto (Ed.), *Prosody and meaning* (pp. 239–270). Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110261790.239

Bolinger, D. (1985). *Intonation and its parts: Melody in spoken English*. Redwood City: Stanford University Press. https://doi.org/10.1515/9781503622906

Bolinger, D. (1989). *Intonation and its uses: Melody in grammar and discourse*. Redwood City: Stanford University Press. https://doi.org/10.1515/9781503623125

Bornkessel-Schlesewsky, I., Sharrad, I., Howlett, C. A., Alday, P. M., Corcoran, A. W., Bellan, V., et al. (2022). Rapid adaptation of predictive models during language comprehension: Aperiodic EEG slope, individual alpha frequency and idea density modulate individual differences in real-time model updating. *Frontiers in Psychology*, *13*, 817516. https://doi.org/10.3389/fpsyg.2022.817516, PubMed: 36092106

Bürkner, P.-C., Gabry, J., Weber, S., Johnson, A., Modrak, M., Badr, H. S., et al. (2023). *brms: Bayesian regression models using 'Stan' (2.19.0)* [Computer software]. https://cran.r-project.org/web/packages/brms/index.html

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

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2009). Sensory processing of linguistic pitch as reflected by the mismatch negativity. *Ear and Hearing*, *30*, 552–558. https://doi.org/10.1097/AUD.0b013e3181a7e1c2, PubMed: 19546808

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*, 2044–2055. https://doi.org/10.1016/j.neuropsychologia.2012.05.004, PubMed: 22595658

Clark, A. (2013). Are we predictive engines? Perils, prospects, and the puzzle of the porous perceiver. *Behavioral and Brain Sciences*, *36*, 233–253. https://doi.org/10.1017/S0140525X12002440, PubMed: 23814868

Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, *1*, 425–452. https://doi.org/10.1515/labphon.2010.022

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009, PubMed: 15102499

Donchin, E. (1981). Presidential address, 1980. Surprise!... Surprise? *Psychophysiology*, *18*, 493–513. https://doi.org/10.1111/j.1469-8986.1981.tb01815.x, PubMed: 7280146

Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., et al. (2009). Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology*, *120*, 1883–1908. https://doi.org/10.1016/j.clinph.2009.07.045, PubMed: 19796989

Emmendorfer, A. K., Correia, J. M., Jansma, B. M., Kotz, S. A., & Bonte, M. (2020). ERP mismatch response to phonological and temporal regularities in speech. *Scientific Reports*, *10*, 9917. https://doi.org/10.1038/s41598-020-66824-x, PubMed: 32555256

Fothergill, A., & Byatt, A. (Directors). (2003). *Deep blue*. BBC Worldwide/Greenlight Media AG.

Friederici, A. D., Wang, Y., Herrmann, C. S., Maess, B., & Oertel, U. (2000). Localization of early syntactic processes in frontal and temporal cortical areas: A magnetoencephalographic study. *Human Brain Mapping*, *11*, 1–11. https://doi.org/10.1002/1097-0193(200009)11:1<1::aid-hbm10>3.0.co;2-b, PubMed: 10997849

Friston, K. (2018). Does predictive coding have a future? *Nature Neuroscience*, *21*, 1019–1021. https://doi.org/10.1038/s41593-018-0200-7, PubMed: 30038278

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., et al. (2004). Hemispheric roles in the perception of speech prosody. *Neuroimage*, *23*, 344–357. https://doi.org/10.1016/j.neuroimage.2004.06.004, PubMed: 15325382

Grice, M. (2022). Autosegmental-metrical phonology—Unpacking the boxes. *Zeitschrift für Sprachwissenschaft*, *41*, 393–411. https://doi.org/10.1515/zfs-2022-2002

Grice, M., Baumann, S., & Benzmüller, 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*, 253–260. https://doi.org/10.1177/00238309211015768, PubMed: 34110260

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., & Elison, T. M. (submitted). Rises on pitch accents and edge tones affect serial recall performance at item and domain levels.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews. Neuroscience*, *8*, 393–402. https://doi.org/10.1038/nrn2113, PubMed: 17431404

Himmelmann, N. P., & Primus, B. (2015). *Prominence beyond prosody—A first approximation* (pp. 38–58). Viterbo: DISUCOM Press.

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

Hsu, Y.-F., Tu, C.-A., Chen, Y., & Liu, H.-M. (2023). The mismatch negativity to abstract relationship of tone pairs is independent of attention. *Scientific Reports*, *13*, 9839. https://doi.org/10.1038/s41598-023-37131-y, PubMed: 37330612

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135. https://doi.org/10.1016/j.brainres.2015.02.014, PubMed: 25708148

Jun, S.-A. (Ed.). (2014). *Prosodic typology II: The phonology of intonation and phrasing*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199567300.001.0001

Kakouros, S., Salminen, N., & Räsänen, O. (2018). Making predictable unpredictable with style - Behavioral and electrophysiological evidence for the critical role of prosodic expectations in the perception of prominence in speech. *Neuropsychologia*, *109*, 181–199. https://doi.org/10.1016/j.neuropsychologia.2017.12.011, PubMed: 29247667

Kember, H., Choi, J., Yu, J., & Cutler, A. (2021). The processing of linguistic prominence. *Language and Speech*, *64*, 413–436. https://doi.org/10.1177/0023830919880217, PubMed: 31631754

Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511808814

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*, 1989–2001. https://doi.org/10.1016/j.jmva.2009.04.008

Li, X., & Chen, Y. (2018). Unattended processing of hierarchical pitch variations in spoken sentences. *Brain and Language*, *183*, 21–31. https://doi.org/10.1016/j.bandl.2018.05.004, PubMed: 29778062

Liu, M., Chen, Y., & Schiller, N. O. (2016). Online processing of tone and intonation in Mandarin: Evidence from ERPs. *Neuropsychologia*, *91*, 307–317. https://doi.org/10.1016/j.neuropsychologia.2016.08.025, PubMed: 27568075

Lorenzen, J., Roessig, S., & Baumann, S. (2022). Information status and tonal context jointly modulate prosodic prominence relations in German. In *Proceedings of the 11th International Conference on Speech Prosody* (pp. 7–11). https://doi.org/10.21437/SpeechProsody.2022-2

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*, 438–445. https://doi.org/10.1016/j.bandc.2011.08.014, PubMed: 21925782

Maess, B., Schröger, E., & Widmann, A. (2016a). High-pass filters and baseline correction in M/EEG analysis. Commentary on: "How inappropriate high-pass filters can produce artefacts and incorrect conclusions in ERP studies of language and cognition". *Journal of Neuroscience Methods*, *266*, 164–165. https://doi.org/10.1016/j.jneumeth.2015.12.003, PubMed: 26774029

Maess, B., Schröger, E., & Widmann, A. (2016b). High-pass filters and baseline correction in M/EEG analysis-continued discussion. *Journal of Neuroscience Methods*, *266*, 171–172. https://doi.org/10.1016/j.jneumeth.2016.01.016, PubMed: 26812439

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*, 201–233. https://doi.org/10.1017/S0140525X00078407

Näätänen, R. (1992). *Attention and brain function*. Lawrence Erlbaum Associates, Inc.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, *38*, 1–21. https://doi.org/10.1111/1469-8986.3810001, PubMed: 11321610

Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*, 313–329. https://doi.org/10.1016/0001-6918(78)90006-9, PubMed: 685709

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. https://doi.org/10.1093/oso/9780198705079.003.0001

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*, 4–22. https://doi.org/10.1111/j.1469-8986.2010.01114.x, PubMed: 20880261

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*, 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026, PubMed: 17931964

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, *125*, 826–859. https://doi.org/10.1037/0033-2909.125.6.826, PubMed: 10589304

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, *9*, 97–113. https://doi.org/10.1016/0028-3932(71)90067-4, PubMed: 5146491

Ou, S.-C., & Guo, Z.-C. (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*, 437–466. https://doi.org/10.1177/0023830919886604, PubMed: 34110259

Paavilainen, P., Arajärvi, P., & Takegata, R. (2007). Preattentive detection of nonsalient contingencies between auditory features. *NeuroReport*, *18*, 159–163. https://doi.org/10.1097/WNR.0b013e328010e2ac, PubMed: 17301682

Peters, J. (2018). Phonological and semantic aspects of German intonation. *Linguistik Online*, *88*, 85–107. https://doi.org/10.13092/lo.88.4191

Polich, J. (1986). Normal variation of P300 from auditory stimuli. *Electroencephalography and Clinical Neurophysiology*, *65*, 236–240. https://doi.org/10.1016/0168-5597(86)90059-6, PubMed: 2420577

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*, 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019, PubMed: 17573239

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25. https://doi.org/10.1080/00335558008248231, PubMed: 7367577

Quasthoff, U., & Richter, M. (2005). Projekt deutscher wortschatz. *Babylonia*, *3*, 33–35.

Ren, G.-Q., Yang, Y., & Li, X. (2009). Early cortical processing of linguistic pitch patterns as revealed by the mismatch negativity. *Neuroscience*, *162*, 87–95. https://doi.org/10.1016/j.neuroscience.2009.04.021, PubMed: 19374939

R Core Team. (2021). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

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*, 66–72. https://doi.org/10.1016/j.neuroimage.2005.01.017, PubMed: 15862206

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*, 135–143. https://doi.org/10.1016/j.brainres.2006.01.043, PubMed: 16487946

Roessig, S., Mücke, D., & Grice, M. (2019). The dynamics of intonation: Categorical and continuous variation in an attractor-based model. *PLoS One*, *14*, e0216859. https://doi.org/10.1371/journal.pone.0216859, PubMed: 31120891

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*, 33–59. https://doi.org/10.1080/23273798.2020.1779324

Röhr, C. T., Savino, M., & Grice, M. (2022). The effect of intonational rises on serial recall in German. In *Proceedings of the 11th International Conference on Speech Prosody* (pp. 759–763). https://doi.org/10.21437/SpeechProsody.2022-154

Savino, M., Winter, B., Bosco, A., & Grice, M. (2020). Intonation does aid serial recall after all. *Psychonomic Bulletin & Review*, *27*, 366–372. https://doi.org/10.3758/s13423-019-01708-4, PubMed: 31974867

Seidl, A., & Johnson, E. K. (2006). Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental Science*, *9*, 565–573. https://doi.org/10.1111/j.1467-7687.2006.00534.x, PubMed: 17059453

Sokolov, E. N. (1963). Higher nervous functions: The orienting reflex. *Annual Review of Physiology*, *25*, 545–580. https://doi.org/10.1146/annurev.ph.25.030163.002553, PubMed: 13977960

Streefkerk, B. M. (2002). *Prominence. Acoustic and lexical/syntactic correlates*. UtrechtLOT. https://dare.uva.nl/search?identifier=44dd581d-7c95-442b-96fe-9a6a1e19c1f3

Sussman, E. S., Chen, S., Sussman-Fort, J., & Dinces, E. (2014). The five myths of MMN: Redefining how to use MMN in basic and clinical research. *Brain Topography*, *27*, 553–564. https://doi.org/10.1007/s10548-013-0326-6, PubMed: 24158725

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*, 268–272. https://doi.org/10.1016/j.neulet.2010.10.035, PubMed: 20970477

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*, 624–628. https://doi.org/10.1097/WNR.0000000000001453, PubMed: 32345849

von Heusinger, K., & Schumacher, P. B. (2019). Discourse prominence: Definition and application. *Journal of Pragmatics*, *154*, 117–127. https://doi.org/10.1016/j.pragma.2019.07.025

Wang, L., Bastiaansen, M., Yang, Y., & Hagoort, P. (2011). The influence of information structure on the depth of semantic processing: How focus and pitch accent determine the size of the N400 effect. *Neuropsychologia*, *49*, 813–820. https://doi.org/10.1016/j.neuropsychologia.2010.12.035, PubMed: 21195102

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (2nd ed.). Cham: Springer. https://doi.org/10.1007/978-3-319-24277-4

Wickham, H., Averick, M., Bryan, J., Chang, W., D'McGowan, L., François, R., et al. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*, 1686. https://doi.org/10.21105/joss.01686

Widmann, A., Schröger, E., & Maess, B. (2015). Digital filter design for electrophysiological data—A practical approach. *Journal of Neuroscience Methods*, *250*, 34–46. https://doi.org/10.1016/j.jneumeth.2014.08.002, PubMed: 25128257

Winkler, I. (2007). Interpreting the mismatch negativity. *Journal of Psychophysiology*, *21*, 147–163. https://doi.org/10.1027/0269-8803.21.34.147

Wolff, S., Schlesewsky, M., Hirotani, M., & Bornkessel-Schlesewsky, I. (2008). The neural mechanisms of word order processing revisited: Electrophysiological evidence from Japanese. *Brain and Language*, *107*, 133–157. https://doi.org/10.1016/j.bandl.2008.06.003, PubMed: 18667231