

# **Rises on Pitch Accents and Edge Tones Affect Serial Recall Performance at Item and Domain levels**

Martine Grice<sup>1(\*)</sup>, Michelina Savino<sup>2,1</sup>, Petra B. Schumacher<sup>3</sup>, Christine T. Röhr<sup>1</sup>, T. Mark Ellison<sup>1</sup>

<sup>1</sup>IfL-Phonetics, University of Cologne, Cologne, GERMANY

<sup>2</sup>Dept of Education, Psychology, Communication, University of Bari, Bari, ITALY

<sup>3</sup>Dept of German Language and Literature I, Linguistics, University of Cologne, Cologne, GERMANY

[martine.grice@uni-koeln.de](mailto:martine.grice@uni-koeln.de), [michelina.savino@uniba.it](mailto:michelina.savino@uniba.it), [petra.schumacher@uni-koeln.de](mailto:petra.schumacher@uni-koeln.de),  
[christine.roehr@uni-koeln.de](mailto:christine.roehr@uni-koeln.de), [tmarkellison@gmail.com](mailto:tmarkellison@gmail.com)

(\*) Corresponding author

Word count: 7776 (excluding title, references and notes/acknowledgements)

## **Abstract**

This paper investigates the effect of intonational rises on attention towards, and ultimately, recall of, medial elements in nine-digit lists in German. Non-final triplets (positions 1, 2, 3 and 4, 5, 6) were produced with either a rise or a fall on digits in positions 3 and 6. Rises led to significantly improved recall over falls. Crucially, the nature and shape of the rise determined the position in which better recall was found. A pitch accent rise on specific digits (at positions 3 and 6) had a local effect on recall of those digits. A boundary rise, marking the end of a triplet, not only boosted recall of the specific digits but also boosted recall of the whole triplet. These results support a prosodic hierarchy in which edge tones are associated with a whole domain (such as an intermediate phrase), rather than simply being placed at its edge, in accounting for the effect on recall of the digits within that domain.

**Keywords:** serial recall, intonation, prosodic hierarchy, rising intonation, prominence, working memory, attention orienting.

## **1. Introduction**

### *1.1. Intonation and attention orienting*

In speech processing, prosodic prominence is crucial for directing listeners' attention to the most important parts of the linguistic message. For instance, in West-Germanic languages important words are typically marked by a specific pitch accent (Baumann, Becker, Grice, & Mücke, 2007; Chen, Den Os, & De Ruiter, 2007; Grice, Ritter, Niemann, & Roettger, 2017; Ito, Speer, & Beckman, 2004; Kohler, 1991; Pierrehumbert & Hirschberg, 1990; Röhr & Baumann, 2010). Some evidence also comes from memory tasks involving the recall of prosodically prominent vs. non-prominent words. For instance, in a recognition memory task on English and Korean (Kember, Choi, Yu, & Cutler, 2019), participants listened to blocks of sentences and at the end of each block, words were presented

to them on a screen and listeners were asked whether they had heard the presented word or not in the preceding block. It was found that English participants were more accurate and faster in recalling prosodically prominent than non-prominent words. In another word recognition memory task on American English sentences containing contrastive pitch accentuation, either a simple H\* peak accent or a rising L+H\* accent, it was found that recognition accuracy increased on accented words that were rising (L+H\*) (Fraundorf, Watson, & Benjamin, 2010). Indications that contrastive rising accents improve word recall were also found for German (Koch & Spalek, 2021)<sup>1</sup>.

The results in Fraundorf et al. (2010) and Koch and Spalek (2021) are compatible with findings in inter-disciplinary research showing that rising F0 contours are perceived as more prominent than falling ones (see e.g. Baumann & Röhr, 2015 for German pitch accent types), since intonational rises attract more attentional resources than falls (Hsu, Evans, & Lee, 2015; Röhr, Brilmayer, Baumann, Grice, & Schumacher, 2020 for German; see also early experimental evidence on infants' attention in Sullivan & Horowitz, 1983). This appears to be relevant in assigning a special contribution to intonational rises in memory performance, also given the role played by attention in fixing stimuli in working memory (Oberauer, 2019).

### *1.2. Intonation in serial recall*

Serial recall tasks require participants to recall a list of items (e.g. digits) in the same order in which they were presented. They are commonly used to assess working memory capacity in both research and clinical settings (Baddeley, Eysenck, & Anderson, 2009; Wechsler, 1987). In these tasks, it is well-documented that the item in the first and last positions are best recalled, owing to primacy and recency effects, respectively (Baddeley et al. 2009). However, when sequences are presented in groups, overall recall accuracy is higher than when sequences are ungrouped: this is the so-called grouping effect, which *also* triggers primacy and recency effects within each group (Ryan, 1969;

---

<sup>1</sup> More specifically, the Koch & Spalek (2021) study shows that a contrastive rising accent on a focused word enhances the recall for contextual alternatives.

Frick, 1989; Baddeley et al. 2009; Crowder & Greene, 2000). In traditional studies on the recall of auditory sequences, prosodic structure has been documented as triggering the grouping effect. These studies show that overall recall accuracy was higher when nine-digit sequences were presented in a grouped sequence with three triplets (e.g. 123-456-789) than in an ungrouped sequence (e.g. 123456789), (see Frankish, 1995 for English; Saito, 1998 for Japanese). However, prosodic structure was frequently cued by the simple insertion of pauses. Crucially, in these studies, the grouping-by-intonation condition was not found to facilitate recall accuracy over and above the grouping-by-pause condition. Note that overall recall accuracy was measured rather than positional effects.

In a recent study involving Italian, Savino et al. (2020) set out to test these results, by partially replicating Frankish's (1995) experiment on English, this time asking Italian participants to recall nine-digit spoken sequences under different prosodic conditions. In the "intonation contour" condition, sequences were characterized by a natural Italian list intonation, the grouping into three being realized by marking items in positions 3 and 6 with a low pitch accent followed by a boundary rise at the end of each of the first two triplets. Moreover, digits in the last position in the sequence (position 9) were marked with a falling boundary tone, as found in natural list endings. In a further condition, items were all realized with a neutral "citation-form" falling intonation, and grouping was achieved by inserting a silent interval between positions 3 and 4, and between positions 6 and 7. These same digits were used for producing the ungrouped lists as controls. Importantly, Savino et al. used concatenated pre-recorded individual digits instead of resynthesized ones, as used by Frankish. This ensured that the Italian stimuli sounded natural, presumably more so than the English stimuli. Savino et al. showed that intonation did in fact enhance overall recall performance, and was consequently better than the condition with only pauses. Serial recall enhancement was found not only in terms of overall recall accuracy, but also and especially for digits positioned at the end of non-final triplets (i.e. positions 3 and 6), which were marked by a boundary rise. They discuss their results partly in terms of the beneficial contribution of the naturalness of the Italian stimuli.

Specifically, they attribute the improvement in recall performance to the perceptual salience (and thus prominence) of the boundary rise.

### *1.3. Accentual/boundary rises and attention*

The perceptual salience of boundary rises needs further discussion, since in languages like Italian, which has both pitch accents and boundary tones, it is generally assumed that pitch accents are the primary encoders of prominence, whereas the main role of boundary tones is that of marking the edges of phrases/domains (Arvaniti, 2020; Ladd, 2008, Grice, 2022). This suggests that rises on a pitch accent orient listeners' attention to the words on which they occur to a greater extent than rises at a boundary. However, the Italian study by Savino et al. (2020) has shown that rising boundaries at the end of non-final triplets enhance recall. This indicates that boundary tones may also attract attention to words, leading to the question as to whether boundary tones can also cue prominence in some way. If they can indeed cue prominence, the question arises as to what the domain of this prominence might be. In current versions of autosegmental-metrical phonology, boundary tones are associated with a prosodic domain such as the intermediate phrase (ip) or the Intonation Phrase (IP). It is conceivable, therefore, that they could cue prominence to the entire domain with which they are associated. That is, a rising boundary tone at the end of a triplet may enhance the prominence of the whole triplet, leading to a less local effect on recall than accentual prominence, which enhances the prominence of a single word (in this case, a single digit).

Figure 1 shows an extract of a prosodic hierarchy (Beckman & Pierrehumbert, 1986; Pierrehumbert and Beckman, 1998; Ladd, 2008; see also Gussenhoven, 2004; Shattuck-Hufnagel & Turk, 1996; Jun, 2005) for a sequence of nine digits (grouped into triplets), with the whole sequence forming an Intonation Phrase (IP) and each triplet an intermediate phrase (ip). Each intermediate phrase has a head association to a Nuclear Pitch Accent (NPA) and each prosodic word (w) is associated with a Pitch Accent (PA). Straight association lines indicate head association. The head of a constituent – in this case the lexically stressed syllable – associates with a pitch accent. Curved

lines, on the other hand, indicate edge association. Here, the association of a tone to an intermediate phrase or Intonation Phrase is primarily manifested in a tone at the edge of that domain (although edge tones have also been shown to affect scaling of tones within their domain of association, Pierrehumbert & Beckman, 1988). The shorter curved lines each originate at a node in the prosodic tree, e.g. the node dominating an intermediate phrase. The Edge Tone is thus associated with the whole domain covered by that node, making it feasible that a rising tone associated with this node may lead to the whole phrase it dominates being enhanced in its prominence. This in turn would attract attention to the whole triplet and would enhance recall on all three digits. It is important to point out that at the outset of autosegmental-metrical phonology, tones were generated by a finite state model and presented as a string, or sequence, without a hierarchical structure, and Nuclear Pitch Accents had no special status over and above Pitch Accents in general (Pierrehumbert, 1980; see discussion in Arvaniti, 2022).

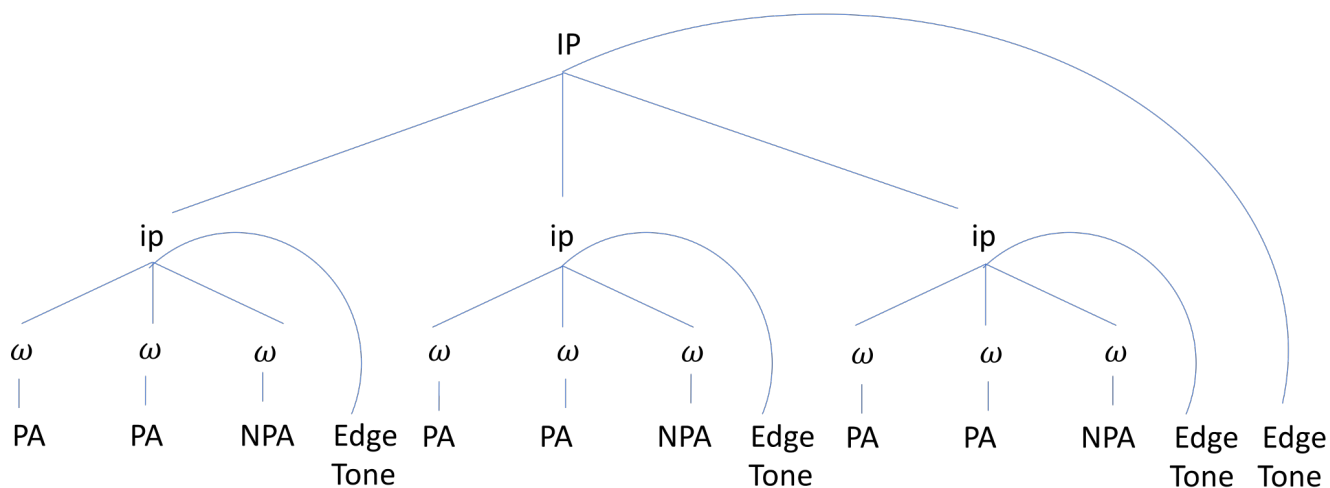


Figure 1. Prosodic hierarchy for a grouped nine-digit list with Pitch Accents (PA) and Edge Tones of the Intonation Phrase (IP) and the intermediate phrases (ip) within it. The head of each ip is a NPA (nuclear pitch accent)

In German, both rising pitch accents and rising boundary tones are typically used by native speakers to recite digit sequences (Baumann & Trouvain, 2001; Peters, 2018). This is because the F0 rising movement at the end of non-final groups signals continuation or non-finality (i.e. that “more is to come” in the list). A falling boundary, by contrast, cues finality (i.e. the end of a sequence, cf. Peters 2018) although falls have also been attested at the end of non-final items (Selting, 2007). However, no previous studies have investigated the effect of boundary rises and falls on prominence.

Recall that, as mentioned in section 1.1 above, it has also been found for German (as well as for other West Germanic languages) that rising pitch accents direct attentional resources to the word upon which they are placed. The question arises as to whether accentual rises and boundary rises can orient listeners’ attention to the same extent or not.

## **2. Motivation for the current study**

In this paper, we aim to further investigate the functional contribution of rising pitch, comparing accentual rises with boundary rises and how they affect recall accuracy. We designed a serial recall task similar to the one used by Savino et al. (2020) with the aim of directly comparing recall performance in these two types of rises. Our aim was to ascertain to what extent rising boundary tones have a similar effect to rising accents in the allocation of attentional resources. In a web-based serial recall task we tested nine-digit sequences with three distinct intonation patterns. When grouped, they were broken into three triplets (positions 123-456-789) with either a rising accent, a rising boundary or a falling boundary in positions 3 and 6. The last triplet of these three conditions was always realized with a falling contour in position 9. A further control condition involved all nine-digit sequences with no sub-groups.

Additionally, since naturalness may have played a role in previous studies, we asked participants to rate the naturalness of the lists they had listened to, so as to ascertain how far our stimuli were appropriate in the different conditions. In particular, if falling intonation generally marks finality, it may be less appropriate list-medially than rising intonation.<sup>1</sup>

### **3. Hypotheses**

Hypotheses HA and HB are concerned with accuracy scores in general, whereas Hypotheses HC and HD are concerned with accuracy at specific positions. HE and HF are concerned with naturalness scores.

HA – Sequences grouped by intonation (Table 1: Ar, Br, and Bf) are recalled better than ungrouped sequences (Table 1: Un), as a consequence of the grouping effect.

HB – Sequences grouped by rising intonation (Ar, Br) are recalled more accurately than sequences grouped by falling intonation (Bf), as rises attract more attentional resources than falls.

HC – Accent rises (Ar) benefit the recall of items on which they occur more than boundary rises (Br) as pitch accents enhance prominence on the words they are associated with.

HD – Boundary rises (Br) benefit the recall of all items in the groups they demarcate, as boundary tones enhance prominence on the whole domain (the triplet) they are associated with.

HE – Ar, Br and Bf are rated as more natural than Un: rises or falls on the final item in a sequence makes the sequence more natural, as grouping is a more natural way of reciting lists.

HF – Ar and Br are rated as more natural than Bf: rises at the end of non-final groups are more natural than falls, as rising F0 cues non-finality whereas falling F0 cues finality.

### **4. Methods and materials**

#### *4.1 Prosodic conditions*

We used nine-digit sequences randomly composed with each digit from 1 to 9. Sequences were produced according to the prosodic conditions in Table 1. They are schematized in Figure 2.



Table 1. Prosodic conditions for the nine-digit sequences

<i>Condition</i>	<i>Description</i>
<b>Ar: Accent rise</b>	Items in position 3 and 6 with a rising nuclear pitch accent, grouping
<b>Br: Boundary rise</b>	Items in position 3 and 6 with a rising boundary tone, grouping
<b>Bf: Boundary fall</b>	Items in position 3 and 6 with a falling boundary tone, grouping
<b>Un: Ungrouped</b> (Control)	Each item realized with a neutral accent with a shallow rise followed by a shallow fall, no grouping

As displayed in Figure 2, all sequences in conditions Ar, Br and Bf end with a falling F0 movement at position 9, cueing the end of the list. Sequences in Ar, Br and Bf differ only in the intonation of the last item in non-final groups (positions 3 and 6). The intonational composition of the last triplet is the same across these conditions. The prosodic structure, which is the same for all lists in the grouped conditions, is shown in Figure 3. The ungrouped condition is one intonation phrase containing one intermediate phrase only, all nine digits being part of that phrase.

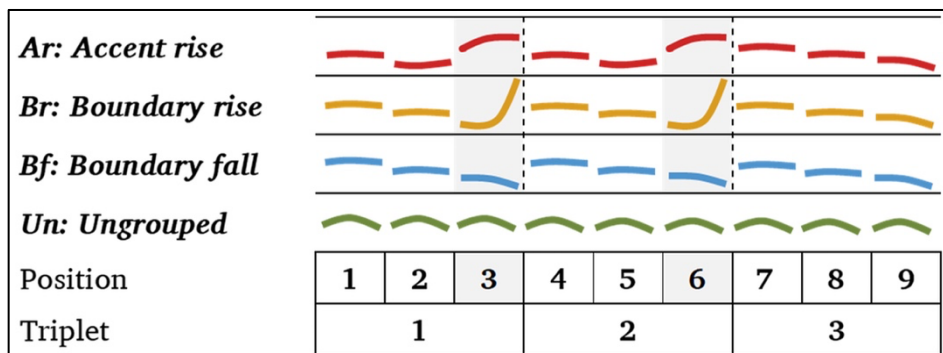


Figure 2. Schematized pitch contours for all four prosodic conditions.

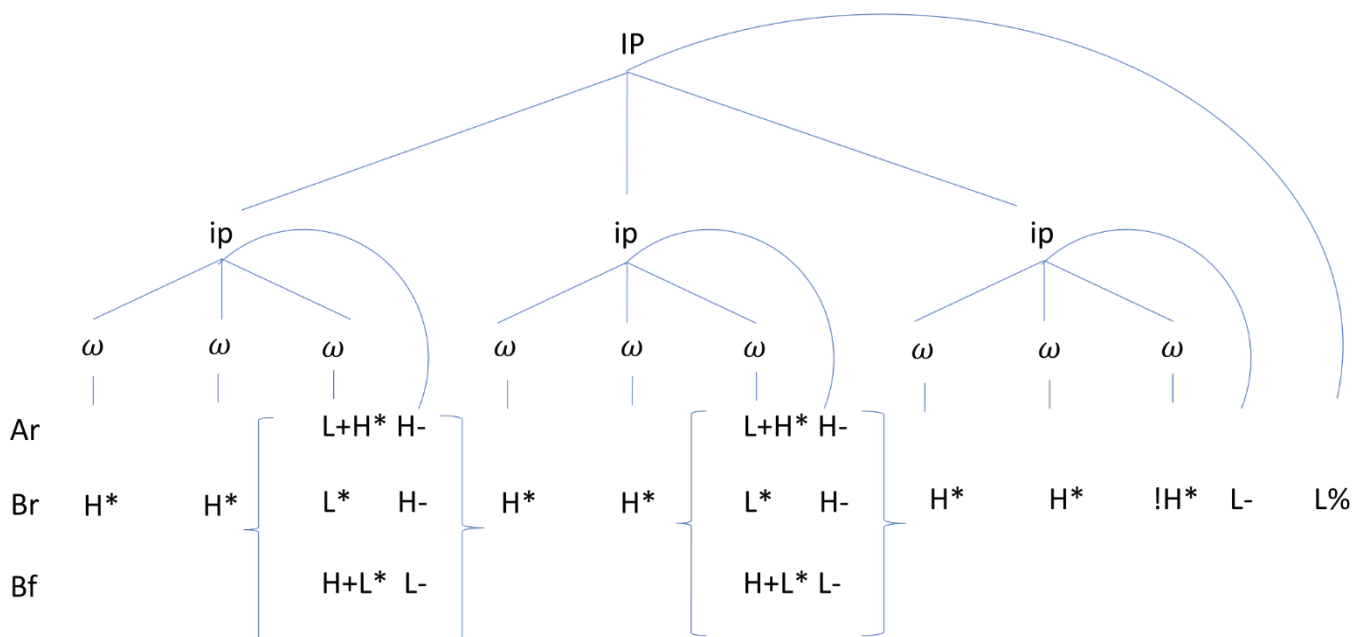


Figure 3. Prosodic structure and tonal representation of the nine-digit lists for the three grouped conditions.

Note that, in contrast to Italian, where digits are all disyllabic words (except for digit three, which is monosyllabic), digits in German are monosyllabic, except for seven (*sieben*) which is disyllabic (1 *eins* [ʔaᶞns], 2 *zwei* [ʔsvaᶞ], 3 *drei* [draᶞ], 4 *vier* [fi:ᶞ], 5 *fünf* [fʏnf], 6 *sechs* [zɛks], 7 *sieben* ['zi:bᶞ], 8 *acht* [ʔaxtʰ], 9 *neun* [nɔᶞn]). However, despite being predominantly monosyllabic, they mostly contain enough sonorant material (although *sechs* and *acht* do have a shorter voiced portion) to allow the intonation contours with distinct shapes to unfold (see Figure 4).

## 4.2 Stimuli preparation

Stimuli were constructed by following the procedure described in Savino et al. (2020). For all digits from 1 to 9, sequences of the same digit in all nine positions were realized in each of the four prosodic conditions. For example, for digit 1 (*eins*) the sequence: *eins eins eins eins eins eins eins eins eins eins* was produced and digitally recorded with Ar, Br, Bf and Un contour types. In this way, all intonational realizations for each position and prosodic condition for each digit were obtained, by taking also into account downtrends in fundamental frequency across stretches of natural speech (Ladd, 1984). These sequences were produced as naturally as possible by a 37-year-old female native German speaker (a trained phonetician) in one recording session, and were recorded at a sampling rate of 44,100 Hz and 16 bit resolution (mono). No adjustments of the recorded sequences were made, except for an equalization of the sound level, using Audacity (Audacity Team, 2021) and the addition of silence before and after digits. We inserted a minimum of 10ms in the digits in all conditions. For the Un condition we needed a minimum of 50ms in order to make the sequences sound natural. We also regularized perceived spacing in the sequences by inserting additional interstitial silences, taking into account features of individual digits, such as the lack of a reflex in the acoustics for closure phases of consonants word initially (e.g. the glottal stop at the beginning of *acht*) and lack of consonants word finally (e.g. *zwei*). Thus, the silent portion of signal between digits was not homogeneous and depended on which digits followed one another. Specific information on how much time was added for each digit is provided in supplementary materials on the OSF repository (<https://osf.io/43eun>).

Digit renditions with added silences for each position under each prosodic condition were saved as individual audio files and used as “building blocks” for creating the nine-digit sequence stimuli. They were constructed by concatenating the individual audio files using Praat (Boersma & Weenink, 2021). The speech waveform and F0 contour of an example of a nine-digit sequence for each of the four prosodic conditions are shown in Figure 4.

Accent rises involve an early rise on the digit, analysed as L + H\*, with some or all of the beginning

of the rise being truncated if the syllable onset was voiceless. This accent rise is followed by a high plateau analysed as H-. Boundary rises, by contrast, have a later rise, with the onset of voicing in the syllable low in a speaker's range. The low pitch is analysed as L\* and the late rise is attributed to an H- boundary tone. The difference is evident between the two rise types, even in monosyllabic digits (eight out of nine of the digits are monosyllabic in German). For example, *eins* (one) in position 3 in the Ar condition has an early rise followed by a high plateau, whereas *drei* (three) in the same position in the Br condition has a low pitch followed by a late rise. Looking at position 6, the digit *sieben* (seven) has predominantly a high plateau, at the onset of voicing the pitch is already high (the rise being truncated), whereas in *neun* (nine) there is a low pitch followed by a late rise. Thus, although Ar and Br conditions both have a rise on the digit, the timing of the rise and the shape of the F0 contour is very different across the two conditions.

Three phoneticians with training in prosodic analysis listened to each of the digits before concatenation, checking perceptual equivalence of the single digits with the same intonation contour across the different numbers (e.g. comparing *drei* with *sieben*) and selecting the most consistent set. Moreover, they also checked each list to guarantee naturalness.

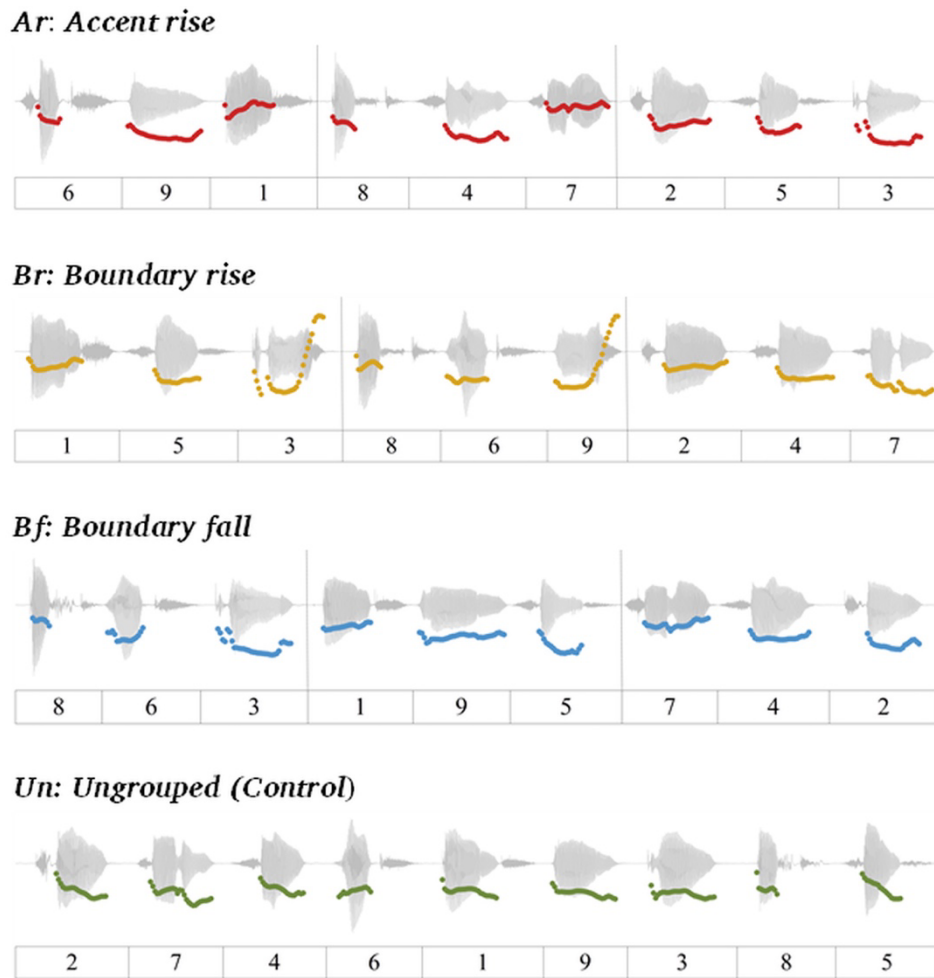


Figure 4. Speech waveforms and F0 contours for a sample nine-digit sequence for each of the four prosodic conditions. F0 range window for all examples: 100-350 Hz.

We produced 17 sequences for each experimental condition following the protocols of previous serial recall studies (Frankish, 1995, Saito, 1998, Savino et al., 2020), so as to facilitate comparisons. This resulted in a total of 68 sequences, including four sequences (one per prosodic condition) to be used as sample items in the task instructions, plus another four (one per prosodic condition) to be used as a training session. The duration of the stimuli sequences averaged 6.2 seconds (SD = 0.2)<sup>2</sup>. The 68 sequences were derived by pseudo-random permutation of the digits 1-9, avoiding (i) two adjacent

<sup>2</sup> The duration of each digit in each condition is discussed and visualised in the Appendix.

digits in ascending or descending order within a sequence and (ii) the same digit in an identical position in consecutive sequences within the same condition.

An additional set of 14 sequences were produced using the digit renditions in the *ungrouped* condition to be used for testing participants with their Digit Span before running the serial recall task. These sequences were constructed according to the WAIS-R Digit Span protocol (Wechsler, 1987).

#### 4.3 Procedure

The task was web-based, implemented with the *SoSci Survey* software and made available to participants at [www.soscisurvey.de](http://www.soscisurvey.de) (Leiner, 2019). Therefore, a stable internet connection, a computer, laptop or tablet, a quiet environment and wearing headphones were specified as technical prerequisites for taking part in the experiment.

In order to ensure comparable hearing conditions for all participants, the task was preceded by a session where they were asked to wear their headphones and carry out a volume calibration test on their devices.

For the main task, participants were instructed to listen to prerecorded nine-digit sequences (containing digits from 1 to 9) and to recall all nine digits of each sequence in the order in which they were presented, by clicking their response on a numeric keypad appearing on the screen. The importance of recalling the nine-digit sequences in the correct order was stressed in the instructions, and participants were also asked not to skip any of the nine digits in their responses, even in case of uncertainty. A counter at the top left above the keypad displayed how many digits had already been entered.

Every sequence was announced by a 263 Hz tone of 890 ms, followed by 500 ms of silence. The numeric keypad was displayed right after a sequence was completely played. Once they had entered nine digits in a trial, participants could proceed to the next one by clicking a “Next” button. They were allowed to listen to each sequence only once.

Stimuli from the same condition were presented as a block (with 16 sequences for each block/condition). The order of presentation of the four blocks was balanced across participants. They were encouraged to take a break at least at the end of each block (1-5 min). Once they had completed the recall of all sequences in a block, participants were additionally asked to rate the naturalness of the previously heard sequences by placing a mark on a visual analogue scale encoding interval data: the left pole labelled as “unnatural” (= 1) and the right pole labelled as “natural” (= 100).

Before starting the experimental task, participants were tested for their Digit Span. As explained above, our DS testing used the pre-recorded concatenated ungrouped monotone digit sequences. The duration of each entire session ranged between 25-50 min (mean time = 37.7, SD = 6.2). As input mode, participants mostly used mouse clicks (69%), in some cases the touchpad (24%), and rarely the touchscreen (7 %).

#### 4.4 Participants

Sixty native speakers of German (27 female, 33 male, aged 18-47 years, mean age = 29.23, SD = 7.75), participated in the experiment, all recruited via the online platform *Prolific* (<https://www.prolific.co/>) on the basis of a number of requirements such as being native monolingual German, having been born and currently living in Germany, and not reporting any auditory, visual or neurological impairment. They all gave informed consent (approved protocol by the ethics committee of the German Linguistic Society #2020-04-200327) and were paid 8 Euros for taking part in the experiment.

One participant had to be excluded from the analysis for not meeting the requirement of being raised monolingually. Submissions from further four participants were also discarded because of their exceptionally high performance (i.e. over 90 % recall accuracy in all conditions, including the control condition), potentially achieved with activities subverting the experimental design, such as writing down the digits whilst listening. The nature of web-based testing (a necessity during the

pandemic) makes such subversion more possible than a test where the participant is supervised in the laboratory, as in all previous work of our own and of other research groups to which we refer. The remaining 55 monolingual German participants (26 female, 29 male) included in the analysis are aged between 18 and 47 years (mean age = 29.22 years, SD = 7.88). Their Digit Span ranged between 4 and 9, with a mean of 6.76 (SD = 1.15).

#### 4.5 Data analysis

We performed non-parametric permutation tests (Berry, Johnston, & Mielke, 2011; Good, 2013; Oden & Wedel, 1975; Pesarin & Salmaso, 2010) to determine likelihoods of the effects of conditioning arising by chance. These were implemented with bespoke parallel software coded in R (R Core Team, 2021). Code and data are available at an OSF repository (<https://osf.io/a85nf/>). These tests explored any effects of the conditions on accuracy across sequences as a whole and in particular positions, but also of naturalness, the latter always a block-level property.

To be more precise, permutation tests were performed on the distributions of correct and false recalls of each stimulus. Items were scored *correct* only if recalled in the same sequence position in which they were presented. A total of 29 700 observations (4 conditions \* 15 sequences \* 9 positions \* 55 participants) were used for analysis. The first of the 16 stimuli sequences in each condition was considered as training and so dropped from analysis. The permutations used for testing hypotheses preserved (where relevant) participants, item-position-within-sequence, and sequence-position-in-block. Thus, when two conditions were being compared, the available permutations were members of a mathematical group of the form  $C_2^{7425}$ , i.e., 7425 independent choices to swap or not swap corresponding values from the two conditions. This group contains around  $10^{2235}$  permutations. However, in practice, there were many fewer substantive permutations that had any effect on evaluation. For example, if a particular participant in sequence 3 at position 4 was always accurate, regardless of condition, then permutations which swapped condition values for this item would not differ in effect from permutations which did not. Permutation between these items can have no effect



on evaluations of accuracy. In the analysis we sampled a random 100 000 permutations from the permutation space in order to estimate how likely any differences in accuracy were to have arrived by chance.

We tested accuracy between conditions for aggregated scores across the full sequence (excluding positions 1 and 9 - since these are dominated by primacy and recency effects), for each position separately, and for three special groups: the final item in the first two triplets (positions 3 and 6), as well as the aggregate scores from the first triplet excluding 1st position (positions 2, 3), and from the second triplet (positions 4, 5, 6). The final triplet is not measured separately as it always carries the same intonation pattern.

For each of the 100 000 permuted data sets, the difference between accuracy measures of the simulated and real data were calculated. The proportion of times in which the simulated data matched or surpassed the actual data set, in accuracy for a given condition across permutations between that condition and another, gives an estimate of the significance of differences between the two conditions.

## **5. Results**

Two measures were examined: *accuracy*, the number of digits recalled in their correct sequence position, and *naturalness*, an assessment of a block of heard sequences with the same intonation pattern, rated on a scale from 0 to 100. These measures were explored in four balanced within-subject conditions: Ar, Br, Bf and Un.

### *5.1 Accuracy - Descriptive statistics*

Three non-condition factors might affect the accuracy of a response. These are: the participant, the position of an item in the sequence, and the position of the sequence in the block. Figure 5 shows the distribution of mean accuracy (aggregated over all other parameters, including the conditions) over these potential influencing factors.

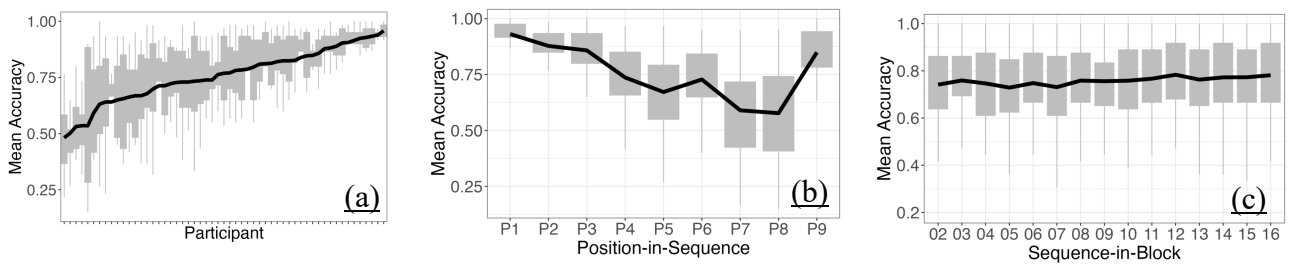


Figure 5. (a) The mean accuracy by participant is plotted with the solid line. For each participant, the accuracies per position-in-sequence were calculated, averaging across sequences in the same block. Participants are ordered here by their mean accuracy, so the monotonicity visible here is an artefact of presentation. The distributions of these accuracies (for the given participant) are shown in the boxplots. (b) The mean accuracy by position-in-sequence, each boxplot showing for that position-in-sequence the distribution of mean accuracy across participants. The solid line shows the standard “bathtub” shape for serial recall curves. (c) The mean accuracy by the position of the sequence within the block, averaged across participants and position in the sequence. There is a slight updrift, but largely constant accuracy. The boxplots show for each sequence-in-block, the distribution of mean accuracies across different participants.

While it is possible to take into account and to model individual differences, and the differing effects of position-in-sequence, this is not the purpose of this paper. The permutations used for assessing the chance likelihood of results do not cross position-in-sequence, sequence-in-block, or participant. So any effect of these will not be confounded with effects of condition.

## 5.2 Accuracy by Condition

Recall accuracy, i.e., the relative number of digits correctly recalled in the position they were presented, is broken down by condition in Figure 6. For readability in this and following graphs, we report on the conditions in decreasing order of accuracy. The overall accuracy is lowest for the

ungrouped, control condition and highest for the three grouped-by-intonation conditions, which are similar in overall performance.

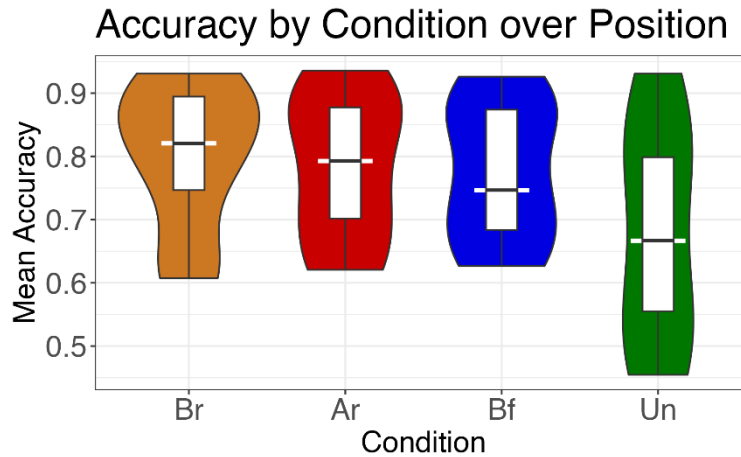


Figure 6. Mean accuracy distributions of results across positions 1-9, grouped by condition. Note the substantial difference between conditions Br, Ar and Bf on the one hand and Un on the other (cf. HA). There are smaller differences between Br, Ar and Bf (cf HB).

Looking at the accuracy of responses at each position in the sequence, and across conditions in Figure 7, we find all conditions show the familiar “bathtub” or “U-curve” shape, reflecting primacy and recency effects. All plots, however, show an uptick at positions 3 and 6 relative to the downdrift trend, even the ungrouped condition Un. This may be the result of an *a priori* grouping bias as well as a highly probable bleed-through from the group conditions to the ungrouped. This uptick leads to the “multiply bowed” appearance. However, the shape of the bowing in the two non-final triplets is not the same across the three intonation contour conditions. In the intonationally grouped conditions, position 6 sees an uptick in accuracy, while the ungrouped condition shows only a reduced rate of fall.

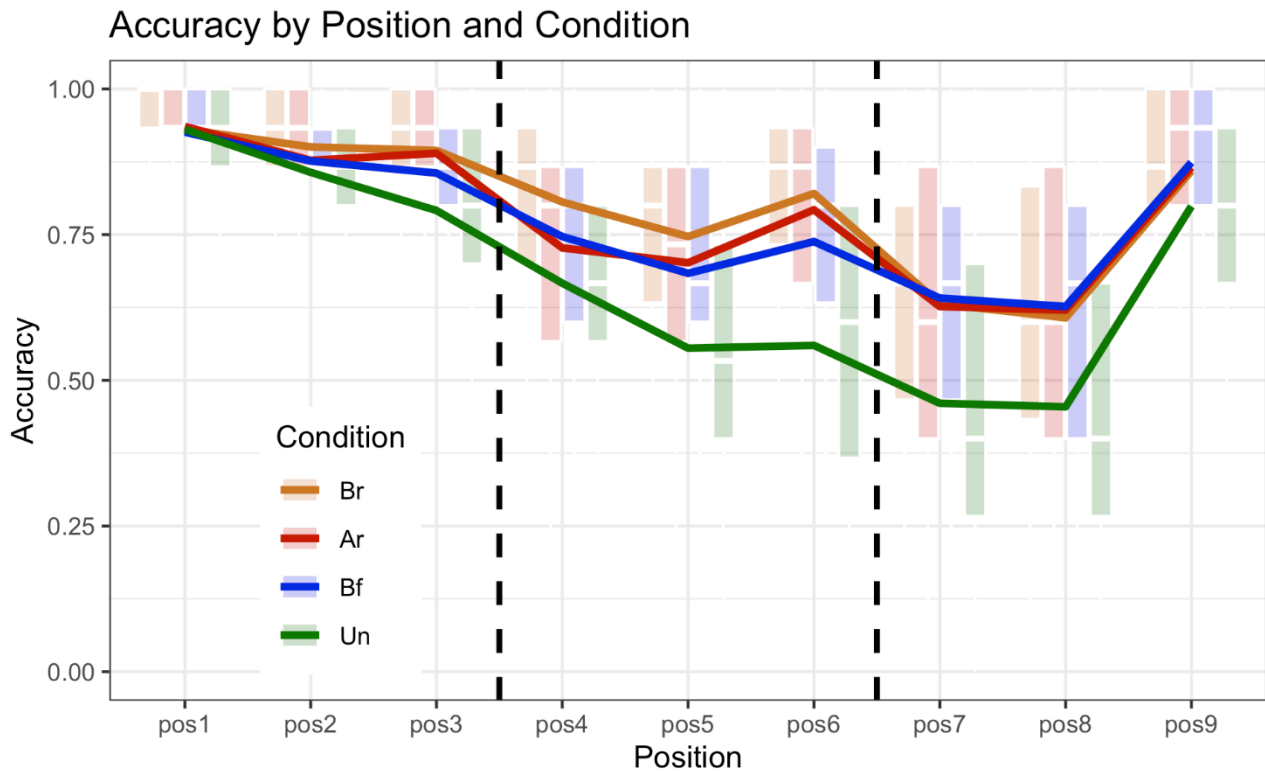


Figure 7. Mean recall accuracy by condition as a fraction of responses. The boxplots at each position show the medians and quartiles of accuracy across permutations affecting conditions only. Notice that mean and median scores often differ, due to asymmetries in accuracy distributions.

### 5.3 Results relating to Hypotheses HA-HB

The pairwise comparisons of the data were performed with a permutation test as described above. For any two conditions being compared, permutations swapped individual item results between conditions, but preserved participant identity, position within the sequence, and position of the sequence within the block. Impacts of these potential factors thus will not differ between permuted results and the actual data.

Differences between simulated and actual data can only arise from effects of condition. The accuracy scores per condition under each permutation (for a given pair of conditions being examined) were calculated, and compared with the score from the actual data set. The fraction  $p$  of permuted data for which the calculated accuracy difference between conditions was higher than the

real data gives an estimate of the likelihood that the experimental accuracy score might have arisen by chance if the two conditions had equal influence on accuracy.

The value  $p$  thus approximates the likelihood of the observed (or greater) level of improved performance arising if the conditions had no effect on the output distribution. Where  $p < 0.05$  we will describe the results as *significant*.

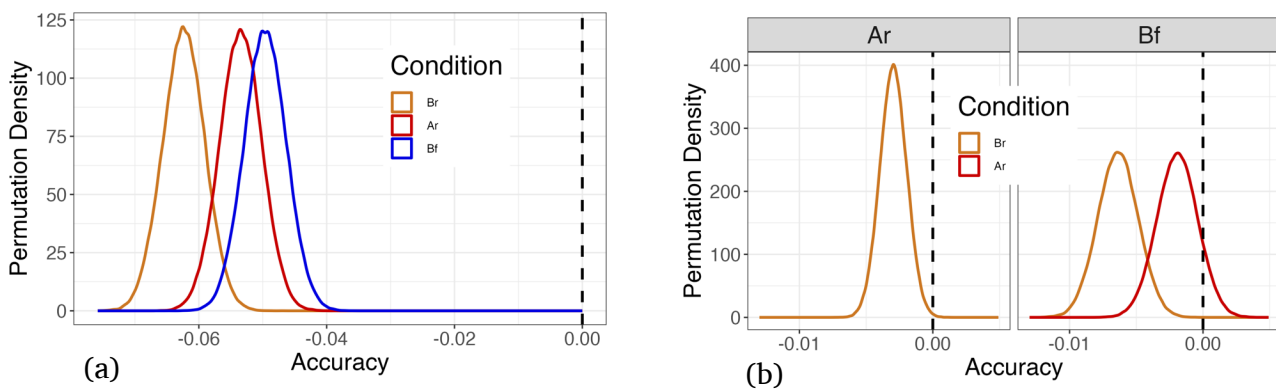


Figure 8. Graph (a) shows the distribution of accuracy difference of permuted to real data where the permutations are with the ungrouped condition Un. The three conditions Br, Ar and Bf show significant differences between real and permuted data, with  $p < 0.001$ . Graph (b) shows the distribution over differences between real and permuted data in conditions in comparisons of Br against Ar and both of these against Bf. The more the dashed vertical line at 0.0 overlaps with the distribution, the more probable the observed difference is by chance, so consequently, the less significant the difference.

Figure 8a shows how far the experimental accuracy for the intonational conditions differs from the distributions of accuracy resulting from permuting these conditions with Un. The fact that 0.0 (the broken vertical line) is a far outlier to those distributions in each case, supports HA: these three conditions result in significantly more accuracy than Un on its own. Figure 8b shows the same comparison for the three conditions Br, Ar and Bf. Note that the broken vertical line is not outside the distribution of permutation results for the Ar~Bf permutation, showing that these could have

resulted from chance. Thus, we do not see a clear overall benefit in the Ar condition, compared to the Bf condition (reflected in the numerical results in Table 2).

These estimated probabilities of achieving these results by chance are shown in Table 2. We see low p-valued significant effects of all prosodically marked conditions against the ungrouped condition. However, while there is a significant difference between the boundary rising and falling conditions (Br and Bf, respectively), we see no significant difference in overall accuracy between pitch rising on accent (Ar) and pitch falling on boundary (Bf).

Table 2. Estimates of the probability of achieving a similar comparison of results as found in the actual data, under the effect of permutation. Only comparisons relevant to HA and HB are shown. The final column shows the significance level (the improbability of achieving this data by chance, with the usual definitions: \* is  $p < 0.05$ , \*\* is  $p < 0.01$ , \*\*\* is  $p < 0.001$ )

C1 < C2		p	
Bf	Br	0.00002	***
Bf	Ar	0.10336	
Un	Br	0.00000	***
Un	Ar	0.00000	***
Un	Bf	0.00000	***

The advantage of the three grouped-by-intonation contour types relative to the control condition is also visible in the heat map in Figure 9. This shows the frequency of confusion between different sequence positions, and we see most accuracy (visible in the colour of the diagonal) for the prosodically grouped conditions, and particularly for Br. The frequency of expressing the item from stimulus position y at position x in recall is shown by the colour in the corresponding cell of the

heatmap, but relative to the likelihood of this combination in the ungrouped condition Un. Where the diagonal is green, this shows accuracy improved relative to Un.

Hypothesis HB expected conditions with rising intonation (Br, Ar) to result in significantly better recall than the condition with only falling intonation (Bf), a result only found for Br. So, hypothesis HB is partially confirmed: only rising boundary intonation offers greater attentional benefits, resulting in greater accuracy, than falling intonation.

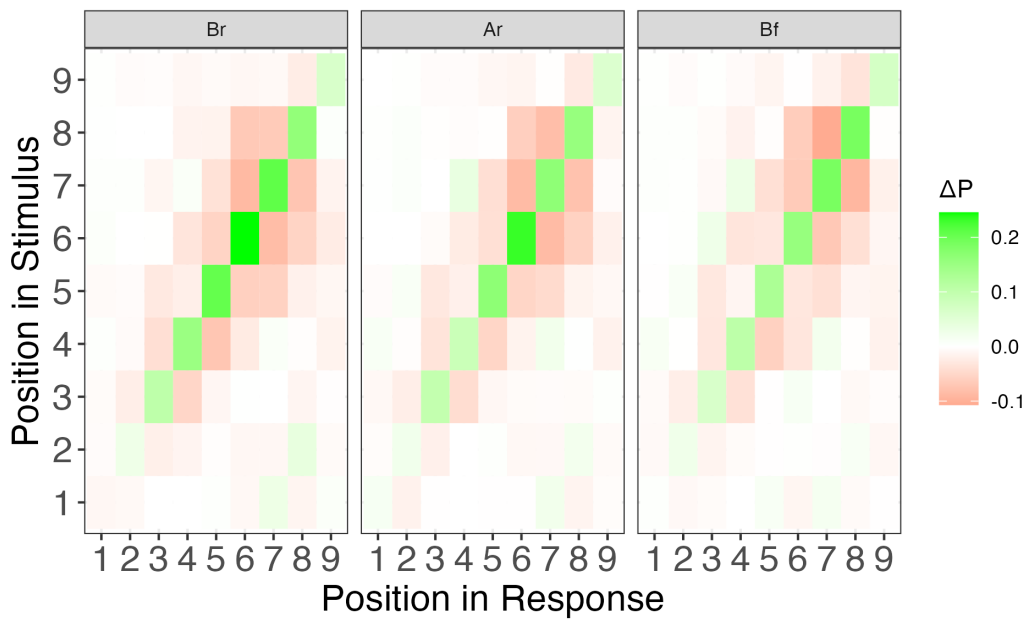


Figure 9. Response rate for various stimulus-response sequence positions relative to control condition (Un). Off-diagonal values indicate errors. Note that digits in all positions other than the first, second and last are recalled better in all three experimental conditions Br, Ar and Bf.

#### 5.4 Results relating to Hypotheses HC-HD

Table 3 shows all significant differences found at particular positions within sequences (see Supplementary materials D5-PValues.csv at <https://osf.io/a85nf/>). For clarity, we address the more straightforward results pertaining to HD first.

Hypothesis HD states that in condition Br, we expect greater accuracy over Ar and Bf in positions 2, 4 and 5, since we hypothesized that a rising boundary tone at positions 3 and 6 would enhance prominence on the whole domain of the triplet. Results in Table 3 indicate that Br does show better

accuracy than Ar or Bf in positions 2, 4 and 5, thus supporting HD. We can see this for each of these positions separately, as well as for the aggregate accuracy when these three positions are treated as a group ( $p < 0.00001$  in comparison with Ar and with Bf).

The primary conclusions are sustained even if we look at the middle triplet (4, 5, 6) alone. Taking this middle triplet as a whole, the effects are clear with Boundary rises leading to better recall of the whole triplet than Accent rises or Boundary falls ( $Br > Ar$ ,  $p < 0.001$ ;  $Br > Bf$ ,  $p < 0.001$ ) and no significant difference between Bf and Ar.

In contrast, HC is only partially supported. Here, we expected a local benefit with rising accents enhancing prominence of a single item. The key positions for condition Ar, namely positions 3 and 6 in which an accent rise is placed, do result in better recall than in condition Bf (in combination,  $p = 0.00014^{***}$ ). However, no significant difference is found in these positions between Ar and Br, or in the aggregate evaluations over the two positions ( $p = 0.92$ ). The picture seen above is also reflected in the differences in accuracy in the heat map in Figure 10.

Table 3. Estimates of the probability of achieving a similar comparison of results per position as found in the actual data, under the effect of permutation.

C1 < C2		P1	P2	P3	P4	P5	P6	P7	P8	P9
Ar	Br	0.0482			0.0000	0.0153				
		6 *			3 ***	9 *				
Bf	Br	0.0346	0.0039	0.0008	0.0010	0.0000				
		1 *	1 **	5 ***	8 **	0 ***				
Bf	Ar			0.0137			0.0019			
				3 *			6 **			



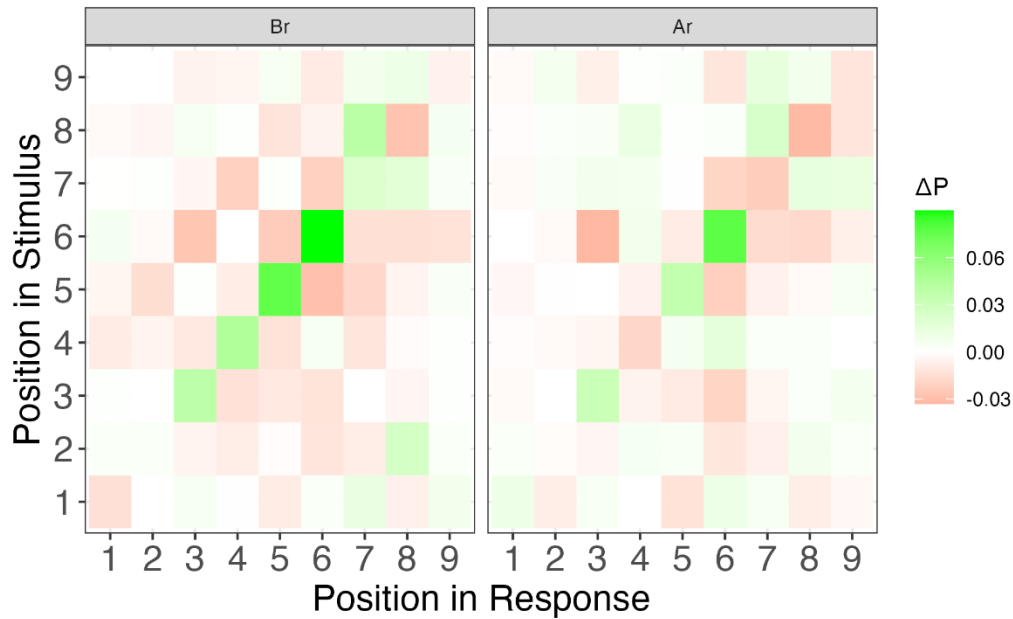


Figure 10. Response rate for various stimulus-answer sequence positions relative to the Bf condition.

Previous work (Ryan, 1969; Frick, 1989; Baddeley et al., 2009; Crowder & Greene, 2000) offers an alternative hypothesis, namely that the initial item in a non-initial group (i.e. an item in position 4 or 7) would receive a primacy advantage in any grouped condition (Ar, Br, Bf) in comparison with the **ungrouped condition (Un)**. We do find an advantage in the grouped conditions, but there is no consistent advantage of one intonation type over the others, as shown in the lack of significant results in position 7 (Ar, Br, Bf > Un). However, in position 4, we also see an advantage of Br over the other two intonation conditions (Ar, Bf). We argue that this is due to a boundary rise on position 6, boosting recall in the whole triplet (4, 5, 6). No similar effect is seen in the final triplet, which lacks a boundary rise (the intonation being constantly falling on the final digit across conditions).

The heat map in Figure 10 shows the difference in recall accuracy for the Br condition (left panel) and the Ar condition (right panel) relative to the Bf condition, with position in the stimulus sequence (y axis) against the position in the recalled sequence (x axis). The greater the difference in diagonal tiles, the higher the accuracy for Ar relative to the Bf condition, and for Br relative to Bf condition. Note improved performance over the whole medial triplet (positions 4, 5, 6) for the Br condition,

whereas improvements only occurred on the last item in the same triplet (position 6) for the Ar condition. Note that in condition Ar, there is less confusion between positions 3 and 6 compared to Bf (the browner squares at positions 3, 6, and 6, 3). At the same time, in condition Br, this lack of confusion extends throughout the triplet, positions 2, 5 then 3, 6 then 4, 7 again showing reduced confusion compared to Bf.

### 5.5 Naturalness (relating to Hypotheses HE and HF)

After each block, participants were asked to rate the sequences they heard for naturalness. Figure 11 shows a violin map for each condition of speaker assessments of naturalness using a visual analogue scale (VAS).

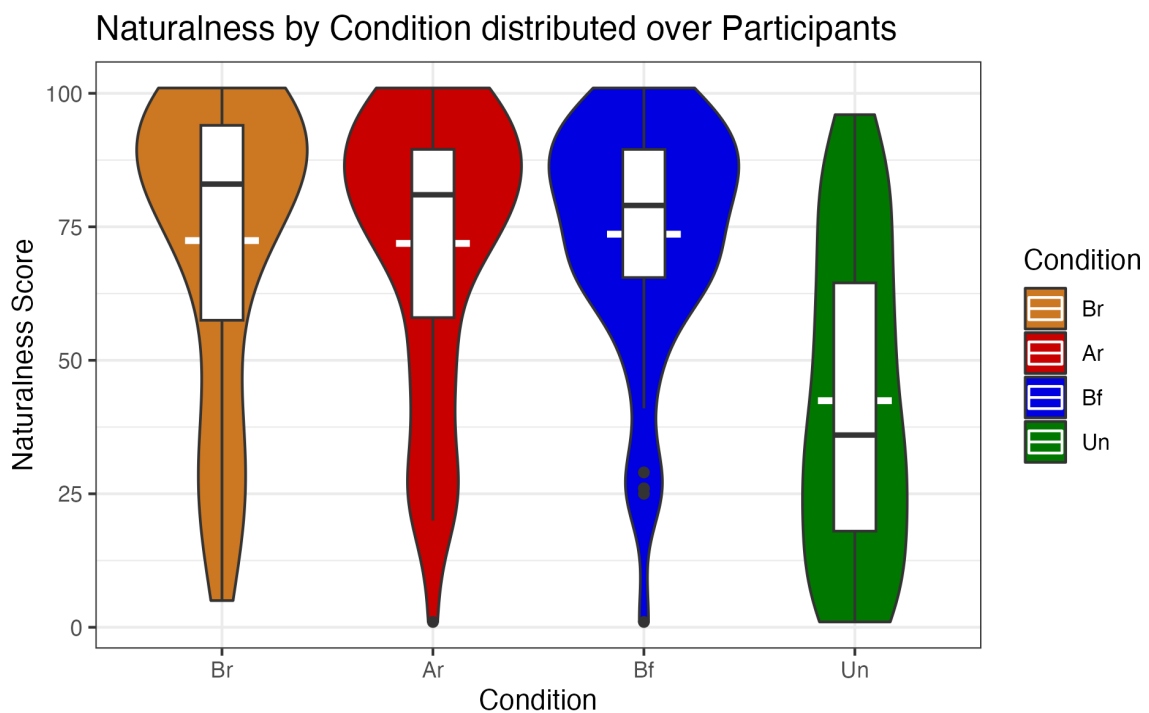


Figure 11. Naturalness ratings by condition, distributed over participants.

Significant differences are shown in Table 4. These confirmed HE, in that intonational grouping, whether rising or falling, accent or boundary, is rated as more natural than the ungrouped control condition. Contrary to predictions from the literature, HF was only partially confirmed, as Br and Ar are not rated as more natural than Bf. In fact, the only significant difference between experimental conditions was that Br was judged significantly more natural than Ar.

Table 4. Likelihood of degrees of difference arising by chance if condition-pair difference does not affect naturalness ratings. The ungrouped condition Un has substantially lower naturalness ratings, a difference extremely unlikely to have resulted from chance. Br with a boundary rise is perceived as significantly more natural than an accent rise.

C1 < C2			p
Ar	Br	0.03969	*
Un	Br	0.00000	***
Un	Ar	0.00000	***
Un	Bf	0.00000	***
Bf	Br	0.05710	
Bf	Ar	0.08224	

## 6. Discussion and Conclusions

In line with previous findings by Savino et al (2020), intonational cues had a positive effect on serial recall. In relation to our hypotheses, HA was confirmed. As predicted by the grouping effect, all three conditions with intonational marking at positions 3 and 6 (**Accentual rise**, **Boundary rise** and **Boundary fall**) led to better recall than the control ungrouped condition (**Ungrouped**).

HB was also confirmed for the two non-final triplets: rising boundary pitch appears to direct attention to items more effectively than falling pitch. However, rising accentual pitch did not lead to greater accuracy than falling pitch, in the overall score.

Results regarding the nature of the tone (whether a rise is associated with a head or an edge of a prosodic domain, i.e. a pitch accent or boundary tone respectively) are related to hypotheses HC and HD. They indicate that accent rises do not boost recall more than boundary rises on the items on which they occur (positions 3 and 6), thus disconfirming HC. As suggested by one of our reviewers, the effect of an accent on memory is not necessarily expected to be local, as found in the Koch & Spalek (2021) study, which showed that a focus accent does not necessarily improve the recall of the accented word itself (but rather the recall of its alternatives). It remains to be investigated if a rising boundary after a focused item aids recall of that item more than a falling boundary or an accent on the focused item.

On the other hand, boundary rises boost recall of nearly all of the groups they demarcate (a boundary rise at position 3 boosts recall of 2; a boundary rise at 6 boosts recall of 4 and 5), confirming HD. Here the boundary rise appears to highlight the whole of the triplet and thus improves recall of all three digits, as compared to the accent rise highlighting the final digits (positions 3 and 6) only. Consequently, accent rises appear to have a more local effect, directing attention to the digit on which they are placed (rather like a laser pointer trained on the one digit), whereas boundary rises tend to have an effect that is distributed over the whole triplet within the series (rather like a wider angled spotlight trained on the entire triplet).

We have thus shown that intonation can serve to highlight one or more items in a list, leading to improved recall, and that the nature of the intonational tones (whether accentual or boundary tones) can determine the locality and extent of such improvements. That is, accent rises serve to highlight the item with which they are associated, in this case the head of an intermediate phrase (here a single digit), whereas boundary rises serve to highlight the whole intermediate phrase with which they are associated (here the entire triplet). This result provides some evidence for the prosodic

hierarchy in which the boundary tone is a property of a whole domain, rather than a sequential view of intonation in which the boundary is simply at the end of a string of tones.

Results also indicate that rising pitch attracts more attention than falling pitch, regardless of phonological status. In addition, this finding for German, together with the previous finding for Italian, point to this effect being cross-linguistic. It is interesting here to point out that, even if both conditions have H-, it is the one with L\* that has a more global effect on recall, not the L + H\*. That is, when the rise is late (Br) the overall amount of maximal pitch on each triplet-final digit is substantially reduced in comparison to when the rise is earlier (where maximal pitch is sustained for longer). It appears not to be the high pitch but rather the rise - a distinction that is not always clear - that is pertinent to group recall.

In more general terms, these findings shed new light on the nature of prosodic prominence and its attention-boosting function. Boundary rises are prominent due to their rising contour and their placement at the edge (the recency effect), resulting in the alignment of prosodic and positional prominence-lending cues (Streefkerk, 2002; Gussenhoven, 2011; see also Himmelmann & Primus, 2015 for other edge placement phenomena). In the present study edge tones draw attention to a larger domain than accentual rises, i.e. they raise prominence globally over a domain. Local prominence (e.g., singling out an item in the current case) and global prominence (e.g., highlighting a prosodic phrase) are thus distinguishable, with the latter enhancing attention and recall across its entire domain. The current findings can be paralleled with research at the level of discourse management that also dissociates cues enhancing local from global discourse prominence (Hinterwimmer, 2019).

At its broadest, the results of the current study show that, given appropriate intonational attributes, prosodic structure is translated into cognitive groupings that can be enhanced as a unit. In everyday speech communication, this grouping, together with the attention-boosting function of rises, means that a particular word or concept that is produced with a rise is paid more attention, is perceived as more prominent, and thus is more likely to be remembered. On a more applied note,

sequences of digits, such as IBAN and orally conveyed confirmation codes could be more efficiently presented with supportive intonation patterns, thus aiding recall and reducing the need for repetition.

As a possible direction for future research, we would expect to see sustained benefits in longer sequences of digits from rising boundary tones in medial positions. This might allow us to resolve questions around internal primacy and recency effects, which, in the present study, are confounded by the finality effect in the last triplet. Another future direction will involve looking into individual differences in the effect of intonation of recall, as it may be the case that some people are less sensitive to the effects of intonation than others.

### ***Notes***

1. A preliminary version of this study using a different methodology for data analysis was presented at Speech Prosody 2022 (Lisbon, May 23-26) and published in the conference proceedings (Röhr, Savino, & Grice, 2022).

### ***Acknowledgements***

Funding acquisition was jointly obtained by Martine Grice and Petra B. Schumacher of the University of Cologne.

### ***Funding***

The study was funded by the Deutsche Forschungsgemeinschaft (German Research Foundation) – Project-ID: 281511265 – SFB 1252 “Prominence in Language” as part of projects A01 “Intonation and attention orienting: Neurophysiological and behavioural correlates” and C09 “Prominence and predictive modelling” at the University of Cologne.

### ***Competing interests***

The authors have no competing interests to declare.

### ***Authors' contributions***

Conceptualization: MG, MS, CTR, PBS. Methodology: MG, MS, CTR. Formal analysis, data curation, visualization: TME. Investigation: MG, MS, CTR. Writing - original draft: MG, Writing - review and editing: MS, CTR, TME, PBS. Supervision and funding acquisition: MG, PBS.

### **References**

- Arvaniti, A. (2020). The Autosegmental-Metrical Theory of Intonational Phonology. In C. Gussenhoven & A. Chen (Eds.), *The Oxford Handbook of Language Prosody*, (pp. 1-20). Oxford: Oxford University Press.
- Arvaniti, A. (2022). The Autosegmental-Metrical model of intonational phonology. In J. Barnes & S. Shattack-Hufnagel (eds.), *Prosodic theory and practice*, (pp.25–63). Cambridge, MA: MIT Press.
- Audacity Team (2021). Audacity(R): Free Audio Editor and Recorder [Computer application]. Version 2.4.2 retrieved February 2, 2021, from <https://audacityteam.org/>
- Baddeley, A., Eysenck M.W., & Anderson, M. C. (2009). *Memory*. London, UK: Psychology Press.
- Baumann, S., Becker, J., Grice, M., & Mücke, D. (2007). Tonal and articulatory marking of focus in German. In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp.1029–1032). <http://www.icphs2007.de/>
- Baumann, S., & Röhr, C. T. (2015). The perceptual prominence of pitch accent types in German. In The Scottish Consortium for ICPHS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS XVIII)* (vol. 298, pp. 1–5). Glasgow, UK: The University of Glasgow. <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0298.pdf>

- Baumann, S., & Trouvain, J. (2001). On the prosody of German telephone numbers. In *Proceedings of the 7th Conference on Speech Communication and Technology* (pp. 557-560). doi: 10.21437/Eurospeech.2001
- Beckman, M. E. & J. B. Pierrehumbert. (1986). Intonational structure in Japanese and English. *Phonology*, 3, 255–309. <https://doi.org/10.1017/S095267570000066X>.
- Berry, K. J., Johnston, J. E., & Mielke Jr, P. W. (2011). Permutation methods. *WIREs Computational Statistics*, 3(6), 527–542. <https://doi.org/10.1002/wics.177>.
- Boersma, P., & Weenink, D. (2021). Praat: Doing phonetics by computer [Computer Program]. Version 6.1.38, retrieved January, 2021, from <https://www.praat.org/>
- Chen, A., Den Os, D., & De Ruiter, J. P. (2007). Pitch accent type matters for online processing of information status: Evidence from natural and synthetic speech. *The Linguistic Review*, 24(2-3), 317–344.
- Crowder, R. G. & Greene, R.L. (2000). Serial Learning. Cognition and Behaviour, In E. Tulving & F. I. M. Craik (Eds.), *The Oxford Handbook of Memory*, (pp.125-136), Oxford: Oxford University Press.
- Frankish, C. (1995). Intonation and auditory grouping in immediate serial recall. *Applied Cognitive Psychology*, 9(7), 5-22.
- Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accent really is. *Journal of Memory and Language*, 63(4), 367-386.
- Frick, R. W. (1989). Explanations of grouping in immediate serial recall. *Memory and Cognition*, 17, 551-562.
- Good, P. (2013). *Permutation tests: A practical guide to resampling methods for testing hypotheses*. Springer Science & Business Media.
- Grice, Martine (2022). Autosegmental-Metrical Phonology - Unpacking the Boxes, *Zeitschrift für Sprachwissenschaft* 41,2: 393-411. <https://doi.org/10.1515/zfs-2022-2002>.



- Grice, M., Ritter, S., Niemann, H., & Roettger, T. B. (2017). Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics*, 64, 90–107.
- Gussenhoven, C. (2011). The sentential prominence in English. In van Oostendorp, M., Ewen, C., Hume, E. & Rice, K. (Eds.). *The Blackwell Companion to Phonology* (pp. 2778-2806). Chichester: Wiley-Blackwell.
- Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press. [10.1017/CBO9780511616983](https://doi.org/10.1017/CBO9780511616983).
- Himmelman, N.P. & Primus, B. (2015). Prominence beyond prosody: a first approximation. In De Dominicis, A. (Ed.) *pS-prominenceS: Prominences in Linguistics. Proceedings of the International Conference* (p. 38-58). Viterbo: Disucom Press.
- Hinterwimmer, S. (2019). Prominent protagonists. *Journal of Pragmatics*, 154, 79-91.
- 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.
- Ito, K., Speer, S., & Beckman, M. (2004). Informational Status and Pitch Accent Distribution in Spontaneous Dialogues in English. In Bel, B. & Marlien, I. (Eds.). *Proceedings of the 2nd International Conference on Speech Prosody*, (pp.279-282). [https://www.isca-speech.org/archive\\_v0/sp2004/sp04\\_279.html](https://www.isca-speech.org/archive_v0/sp2004/sp04_279.html)
- Jun, S.-A. (Ed.). (2005). *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press.
- Kember, H., Choi, J., Yu, J., & Cutler, A. (2019). The processing of linguistic prominence. *Language and Speech*, 64(2), 413-436.
- Koch, X., & Spalek, K. (2021). Contrastive intonation effects on word recall for information-structural alternatives across the sexes. *Memory and Cognition*, 49, 1312-1333.

- Kohler, K. (1991). Terminal intonation patterns in single-accent utterances of German: Phonetics, phonology and semantics. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)*, 25, 115–185.
- Ladd, D. R. (1984). Declination: A review and some hypotheses. *Phonology*, 1, 53-74.
- Ladd, D. R. (2008). *Intonational Phonology* (2nd ed.). Cambridge: Cambridge University Press.
- Leiner, D. J. (2019). SoSci Survey (Version 3.1.06) [Computer software], <https://www.soscisurvey.de>
- Oberauer, K. (2019). Working memory and attention. A conceptual analysis and review. *Journal of Cognition*, 2(1), 1-23.
- Oden, A., & Wedel, H. (1975). Arguments for Fisher's Permutation Test. *The Annals of Statistics*, 3(2), 518–520.
- Pesarin, F., & Salmaso, L. (2010). The permutation testing approach: a review. *Statistica*, 70(4), 481-509.
- Peters, J. (2018). Phonological and semantic aspects of German intonation. *Linguistik Online*, 88(1).
- Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* (Doctoral dissertation, Massachusetts Institute of Technology).
- Pierrehumbert, J. B. & M. E. Beckman. (1988). *Japanese Tone Structure*. Cambridge, Mass.: MIT Press.
- Pierrehumbert, J. B., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollak (Eds.), *Intentions in Communication*, (pp. 271–311). Cambridge: MIT Press.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL <https://www.R-project.org>
- Röhr, C. T., & Baumann, S. (2010). Prosodic marking of information status in German. In *Proceedings of Speech Prosody 2010*, paper 109. [https://www.isca-speech.org/archive\\_v0/sp2010/sp10\\_019.html](https://www.isca-speech.org/archive_v0/sp2010/sp10_019.html)

- Röhr, C. T., Brilmayer, I., Baumann, S., Grice, M., & Schumacher, P. B. (2020). Signal-driven and expectation-driven processing of accent types. *Language, Cognition & Neuroscience*, 36(1), 33-59.
- Röhr, C. T., Savino, M., & Grice, M. (2022). The effects of intonational rises in serial recall in German. In *Proceedings of Speech Prosody 2022*, (pp.759-763), doi: 10.21437/SpeechProsody.2022-154.
- Ryan, J. (1969). Grouping and short-term memory: Different means and patterns of grouping. *Quarterly Journal of Experimental Psychology*, 21, 137–147.
- Saito, S. (1998). Effects of articulatory suppression on immediate serial recall of temporarily grouped and intonated lists. *Psychologia*, 41(2), 95-101.
- Savino, M., Winter, B., Bosco, A., & Grice, M. (2020). Intonation does aid serial recall after all. *Psychonomic Bulletin and Review*, 27, 366-372.
- Selting, M. (2007). Lists as embedded structures and the prosody of list construction as an interactional resource. *Journal of Pragmatics*, 39(3), 483-526.
- Shattuck-Hufnagel, S. & Turk, A.E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25, 193–247.  
<https://doi.org/10.1007/BF01708572>
- Streefkerk, B. (2002). *Prominence. Acoustic and lexical/syntactic correlates*. Utrecht: LOT.
- Sullivan, J. W. & Horowitz, F. D. (1983). The Effects of intonation on infant attention: the role of the rising intonation contour. *Journal of Child Development*, 10, 521-534.
- Wechsler, D. (1987). *Manual for WAIS-R*. New York, NY: Psychological.

## Appendix

As pointed out by an anonymous reviewer, it could have been the case that the stimuli had longer durations in one of the four conditions (e.g. in one of the conditions with a rising intonation at the end of the first and second triplets). If this were the case, it could have affected recall accuracy. However, no such pattern in durations is evident from Figure 12, which shows a heat map of durations by condition and position in sequence. Interestingly, there is no evidence for longer durations in final positions in the triplets (positions 3 and 6) or in the whole sequence (position 9) as compared to other positions. This is because the individual digits differ considerably in their inherent durations, regardless of prosodic contour. Note that although the digit 7 *sieben* ['zi:bɪm] is disyllabic, it is relatively short, owing to reduction of the second syllable.

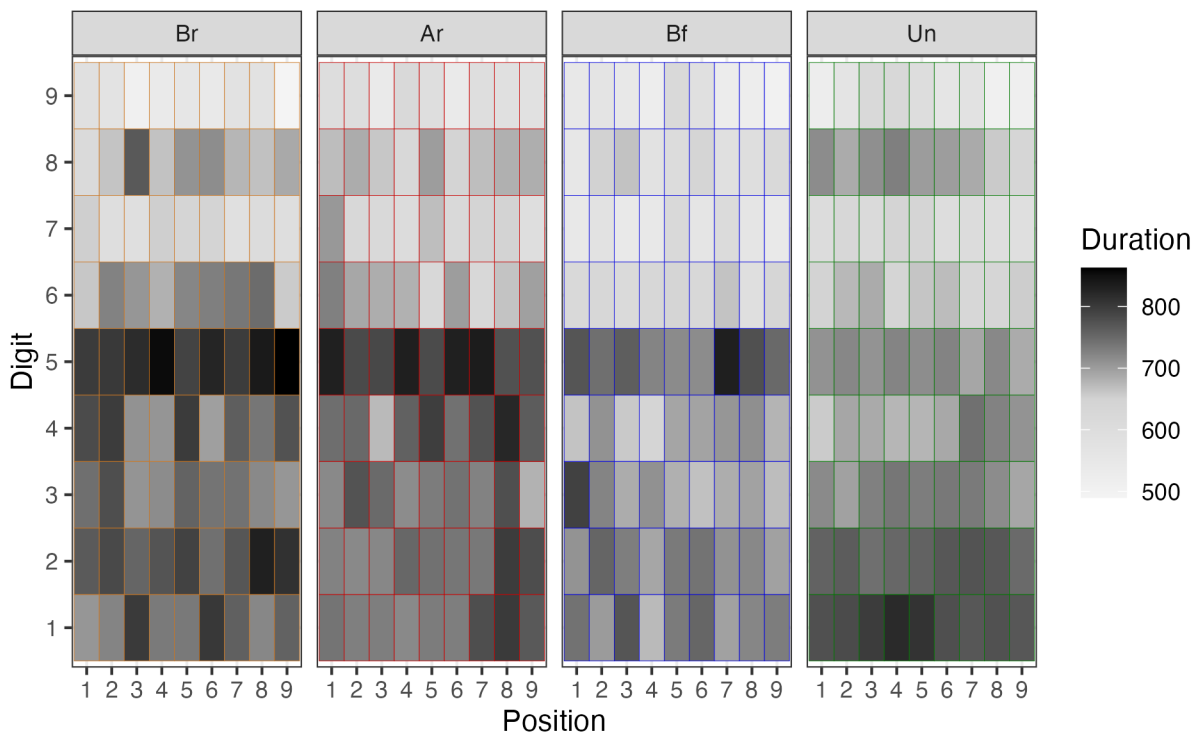


Figure 12 Heat map of durations for each of the conditions in a separate panel (Ar, Br, Bf and Un), with position in sequence on the x-axis and digit on the y-axis. Duration indicated by darkness of each tile (darker = longer, see scale to the right of the heat map).