

Entrainment to speech prosody influences subsequent sentence comprehension

Yulia Lamekina & Lars Meyer

To cite this article: Yulia Lamekina & Lars Meyer (2022): Entrainment to speech prosody influences subsequent sentence comprehension, *Language, Cognition and Neuroscience*, DOI: [10.1080/23273798.2022.2107689](https://doi.org/10.1080/23273798.2022.2107689)

To link to this article: <https://doi.org/10.1080/23273798.2022.2107689>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 11 Aug 2022.



[Submit your article to this journal](#)



Article views: 393



[View related articles](#)



[View Crossmark data](#)

Entrainment to speech prosody influences subsequent sentence comprehension

Yulia Lamekina ^a and Lars Meyer^{a,b}

^aMax Planck Research Group Language Cycles, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany; ^bClinic for Phoniatrics and Pedaudiology, University Hospital Münster, Münster, Germany

ABSTRACT

Speech processing is subserved by neural oscillations. Through a mechanism termed entrainment, oscillations can maintain speech rhythms beyond speech offset. We here tested whether entrainment affects higher-level language comprehension. We conducted four online experiments on 80 participants each. Our paradigm combined acoustic entrainment to repetitive prosodic contours with subsequent visual presentation of ambiguous target sentences (e.g. “Max sees Tom and Karl laughs”). We aimed to elicit faulty segmentations through the duration of the preceding contour (e.g. the segment “Max sees Tom and Karl” leads to an error at “laughs”). Across experiments, self-paced reading data showed that participants employed the duration of the initial prosodic contour to predict the duration of the upcoming segments. Prosody entrainment may thus serve a predictive function during language comprehension, not only helping the reader to segment the current speech input, but also inducing temporal predictions about upcoming segments.

ARTICLE HISTORY

Received 29 October 2021
Accepted 19 July 2022

KEYWORDS

Auditory perception;
language; rhythm; prosody;
psycholinguistics



Introduction


Prosody is critical to sentence comprehension. Pauses, pitch modulations, and duration changes constitute intonational phrase boundaries (IPBs) that delineate multi-word constituents (Frazier et al., 2006). In auditory comprehension, IPBs trigger the segmentation of speech into multi-word units. During reading, where acoustic cues are unavailable, listeners actively construct implicit prosody that facilitates comprehension (Breen, 2014; Breen et al., 2016; Fodor, 2002).

The processing of IPBs and the construction of implicit prosody depend on prosodic context. IPBs are not processed in terms of their absolute acoustic magnitude, but relative to the magnitude of preceding IPBs (Clifton et al., 2002; Snedeker & Casserly, 2010). Even more distant prosodic cues can influence the prediction of upcoming material, such that prosody at sentence onset affects subsequent segmentation and word recognition (Brown et al., 2011; Dilley & McAuley, 2008). Context effects have also been reported across domains; for instance, IPBs can prime segmentation of subsequent visual sentences (Steinhauer & Friederici, 2001). By priming, we mean a phenomenon whereby exposure to a preceding stimulus influences a response

to a subsequent stimulus, without conscious guidance or intention (Branigan, 2007; Pickering & Ferreira, 2008). In the Steinhauer & Friederici (2001) study, structural priming can be considered an explanation of the obtained results – that is, exposure to preceding prosodic contours can pre-activate representations of upcoming structures, which would facilitate the processing of upcoming visual sentences.

Instead of priming, distant context effects might be elicited by rhythmicity of the prior prosodic context. Periodic amplitude- or frequency-modulated sounds at a given frequency improve the auditory detection of subsequent targets that arrive at the same frequency (Henry & Obleser, 2012; Hickok, Farahbod, & Saberi 2015). Likewise, it has been shown that during speech processing, the syllable rate of a lead-in sentence can affect detection of subsequent target syllables: After a fast-rate sentence, subjects overhear short target syllables that they do perceive when the lead-in sentence is presented at a slow rate (Bosker, 2017; Dilley & Pitt, 2010). In neuroscience, such transfer effects are attributed to an electrophysiological mechanism termed *entrainment*, by which neural oscillations inherit a stimulation rhythm to continue after stimulus offset (Kösem et al., 2018; Luo &

CONTACT Yulia Lamekina  lamekina@cbs.mpg.de  Max Planck Research Group Language Cycles, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig 04103, Germany

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/23273798.2022.2107689>

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Poeppel, 2007). Importantly for the current study, it has been suggested that entrainment can also be triggered by prosodic contours, dominated by modulations of pitch and/or amplitude at frequencies below 4 Hertz (Bourguignon et al., 2013; Mai et al., 2016; for review, see Meyer, 2018).

The usage of the term *entrainment* in auditory and cognitive neuroscience must be distinguished from its usage in pragmatics. Weidman et al. (2016) uses the term to denote a “general tendency of two speakers to demonstrate similarity in aspects of their speech over the course of a conversation” (cf. Lehnert-LeHouillier et al., 2020; Edlund, 2011). This definition applies both to lexical units (Brennan, 1996) and features of intensity, pitch, speech rate, and voice quality. Some of these features might be rhythmic enough to trigger entrainment of electrophysiological oscillators, while others are not. In the current study, we refer only to rhythmic features that mark IPBs, relying on recent corpus evidence for a periodicity of prosodic boundaries (Inbar et al., 2020; Stehwiën & Meyer, 2022).

Our series of four experiments combined an initial rhythm of a prosodic contour repeated three times with a subsequent visual target sentence. In Experiment 1, the frequency of the prosodic rhythm was hypothesised to drive participants into a specific segmentation option when processing a target sentence that contained a syntactic ambiguity (i.e. coordination ambiguity; e.g. *Max sees Tom and Karl laughs.*; Hoeks et al., 2002). Results showed that indeed, the initial contour affected downstream reading times. To test whether this apparent entrainment effect would generalise, Experiment 2 combined the same trial structure with shorter, non-ambiguous target sentences. While still significant, the original transfer effect was weakened, suggesting that the duration of the target sentence is critical for the strength of the entrainment effect. This intuition was further confirmed by experiments 3 and 4, which presented progressively longer target sentences.

Methods (experiments 1–4)

Participants

Every experiment was run on 80 participants (German native speakers, right-handed, age range = 18–35 years; for the demographic data of participants in every experiment see Appendix 1). In order to determine the sample size, we conducted a power analysis using G*Power 3 software (Faul et al., 2007) based on effect sizes obtained from previous research with a similar paradigm (Dilley & Pitt, 2010). Our estimate yielded a sample size of 32;

however, we adopted a more conservative sample of 80 to ensure reproducibility of the effects (Aarts et al., 2015). Participants had normal or corrected-to-normal vision and no reported history of neurological or hearing disorders. They were recruited through the online platform Prolific (www.prolific.co); participation was reimbursed with £ 6 per hour. All participants were naive as to the purpose of the study. Written informed consent was obtained prior to the experiment. The study conformed to the guidelines of the Declaration of Helsinki and was approved by the local ethics committee of the University of Leipzig, DE.

Stimuli

In order to investigate the influence of prosodic entrainment on subsequent sentence comprehension, our experiments began with initial auditory presentation of a prosodic contour repeated three times. This prosodic contour always belonged to one of the conditions – SLOW or FAST – characterised by different length. Prosodic contour exposure was followed by visual word-by-word presentation of a target sentence. Contour and sentence durations varied across the experiments. Critically, in every experiment the presentation of the first several visual words was adjusted to contour duration, whereby word presentation rate was calculated from the same auditory sentences from which the contour was extracted. Minor rate changes across experiments result from different sentence lengths across experiments. Presentation of the last words of the sentence, however, was self-paced – that is, the words would remain on screen until participants would push a button (cf. Figure 1).

It was important to match the auditory and visual presentations exactly, so that the critical word (e.g. third noun (N3) in Experiment 1), which was already self-paced, fell potentially within the time interval spanned by one SLOW contour, but outside of FAST contour. That is, if the contours elicited subsequent transfer effects, the reaction time at this word would try to match the timing pre-imposed by the prosodic entrainment – namely, in case of the SLOW condition, the participant would try to fit the critical word into the time interval, defined by the SLOW contour; therefore, the reaction time would be shorter than that in the FAST condition. This allowed us to interpret the self-paced reaction time at this critical word as a measure of the transfer effect of the preceding contours on the sentence comprehension, potentially via entrainment.

After sentence presentation and a delay, some trials were followed by comprehension questions (see Procedure). Transfer effect was inferred from self-paced

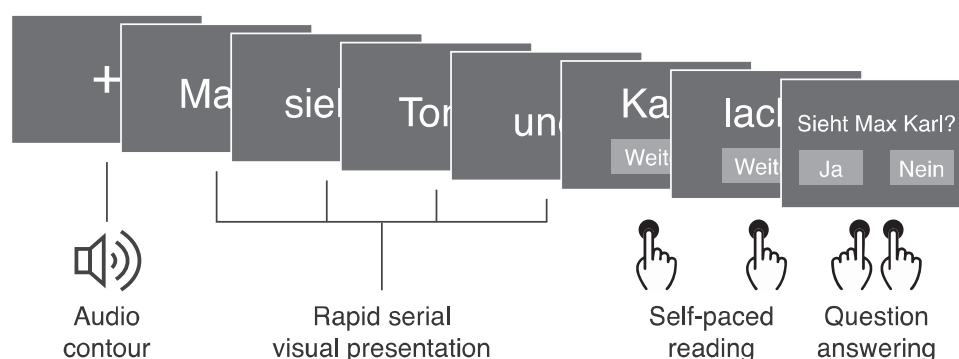


Figure 1. Procedure (all experiments). Participants first listen to the audio contour. Then the sentence is presented word-by-word, apart from the last one/two words, which are self-paced. Finally, in 75% of the trials, a comprehension question follows.

reading (SPR) times at critical words, as well as from reaction times (RTs) and accuracy on comprehension questions. An overview of the stimuli for the 4 experiments is reproduced in Table 1 (self-paced words are indicated in bold; arrows mark the ends of the contours if potentially imposed on the sentence timing). Further, we describe the stimuli for every experiment in more detail.

Experiment 1

Two TYPEs of German target sentences were used (see (1)–(4) for English translations). In the first sentence TYPE, LONG sentences (1), *Karl* is the subject of the verb *laughs*, whereas *Tom* is the object of *sees*. However, if a participant would implicitly construct an IPB after *Karl*, this would trigger the wrong interpretation (2) where both *Tom and Karl* together are the object. This error would be recognised at *laughs*. A similar effect could be expected in the second type of target stimuli, SHORT sentences (3) and (4):

- (1) LONG (correct): [Max_{N1} sees $_{V1}$ Tom_{N2}] [and Karl_{N3} laughs $_{V2}$].
- (2) LONG (error; verb): [Max_{N1} sees $_{V1}$ Tom_{N2} and Karl_{N3}] laughs $_{V2}$.
- (3) SHORT (correct): [Max_{N1} sees Tom_{N2} and Karl_{N3}].
- (4) SHORT (error; no verb): [Max_{N1} sees Tom_{N2}] [and Karl_{N3} \emptyset].

For construction, we used 32 monosyllabic first names of 3–6 characters to balance word-by-word presentation (New et al., 2006). Noun frequencies were normally distributed (Heister et al., 2011). Since male and female first names differ in length, we used male first names only. We also selected 75 transitive and 75 intransitive German verbs in 3rd person singular present tense. Length was matched (1–2 syllables, 5–8 characters). Verb frequencies were also normally distributed. Pairs of transitive and intransitive verbs were made based on

semantic fit (e.g. *expect–come*, *wake up–sleep*). Combination of verb pairs and names yielded 6,000 sentences. A different name triplet was used for each of these. Name triplets were selected to not contain similar-sounding names (e.g. *Frank* and *Franz*).

To elicit the two segmentation options, prosodic contours of two different SPEED values were presented before the target sentences. Contours were made by averaging the pitch tracks of the visual sentences, which were stripped off synthetic recordings (Oord et al., 2016) in Praat (Boersma & van Heuven, 2001). We used a female voice (minimum pitch = 116 Hz, maximum pitch = 267 Hz, average pitch = 191.5 Hz) because of its broad pitch range and high variability. The two entrainment SPEED values were SLOW (based on the 3,000 SHORT sentences; e.g. *Max sees Tom and Karl*) and FAST, based on 3,000 additional sentences (e.g. *Max sees Tom*). Thus, in LONG sentences, the SLOW prosodic contour matched the duration of *Max sees Tom and Karl*, aimed at eliciting an error at *laughs* (2). In contrast, the FAST contour matched the duration of *Max sees Tom*, aimed at eliciting the correct segmentation (1). For SHORT sentences, on the contrary, correct segmentations (3) were expected in the case of SLOW entrainment, but errors (4) were expected in the case of FAST entrainment (Figure 2).

For averaging, contour durations were adjusted to the average duration of the respective sentence recordings (SLOW: 1,570 ms; FAST: 942 ms). For full delexicalisation, PURR (Prosody Unveiling through Restricted Representation) method was used (Sonntag & Portele, 1998). This pipeline was recommended in previous studies for constructing a delexicalised prosodic contour (Steinhauer & Friederici, 2001). The method involves extracting the pitch values from the original contour and constructing a sound by adding a sine wave at pitch, its second harmonic of 1/4 of the amplitude, and its third harmonic of 1/16 of the amplitude (suggested by Klasmeyer, 1997). Therefore, out of the original spectral characteristics of the speech signal, only the pitch modulations are retained,

Table 1. Overview of stimuli used in 4 experiments.

Experiment	Factor	Level	Example stimulus												
1	End of contour	SLOW FAST	↓												
	Target sentence	LONG SHORT	Max	sees	Tom	and	Karl							laughs.	
2	End of contour	SLOW FAST	↓												
	Target sentence	LONG SHORT	Max	sees	Tom	and	Karl.								
3	End of contour	SLOW FAST	↓												
	Target sentence	LONG SHORT	Max	Scholz	sees	Tom	Schmidt	and	Karl	Weiss					laughs.
4	End of contour	SLOW FAST	↓												
	Target sentence	LONG SHORT	Max	Scholz	sees	in	the	city	Tom	Schmidt	and	Karl	Weiss	laughs.	
			Max	Scholz	sees	in	the	city	Tom	Schmidt	and	Karl	Karl	Weiss	Weiss.

which permits to disentangle prosodic modulations from other speech components. Prosodic pitch has proved to provide a substantial contribution to entrainment separately from other acoustic and phonetic features (Teoh et al., 2019). PURR has been tested extensively and compared to other methods, proving over a variety of experiments to have the best functionality and acceptability (listeners recognising the signal as coming from natural human speech) for speech delexicalisation (Kotz et al., 2003; Meyer et al., 2002; Pannekamp et al., 2005). Contours were further normalised to 65 dB and lowered in pitch by 55 Hz to ensure a comfortable hearing level.

Average word duration for timed visual presentation calculated from the synthesised contours length was 314 ms. Presentation of the third noun (N3; e.g. *Karl*) and the second verb (V2; e.g. *laughs*; LONG sentences only) was self-paced; these words stayed on the screen until the participant pressed a button.

The 6,000 sentences and contours were combined into 20 experimental lists of 300 trials each. Within list, every verb pair was used 4 times, once within each

condition (i.e. SHORT–FAST, SHORT–SLOW, LONG–FAST, and LONG–SLOW). Pairs and conditions did not repeat across subsequent trials. We disallowed adjacent name triplets with identical or similar names. Identical triplets did not repeat within list.

Experiment 2

Our second experiment tested the entrainment hypothesis on short non-ambiguous sentences (see Results Experiment 1). Prosodic contours and average visual word duration were identical to Experiment 1. The two conditions were LONG (e.g. *Max sees Tom and Karl*) and SHORT (e.g. *Max sees Tom*). In the SHORT condition, the final word (second noun (N2); e.g. *Tom*) was self-paced; in the LONG condition, the three final words (N2, “und” (U), and third noun (N3); e.g. *Tom, und, and Karl*) were self-paced (see Figure 3). List generation was identical to Experiment 1.

Experiment 3

Our third experiment tested the entrainment hypothesis on longer sentences (see Results Experiment 2).

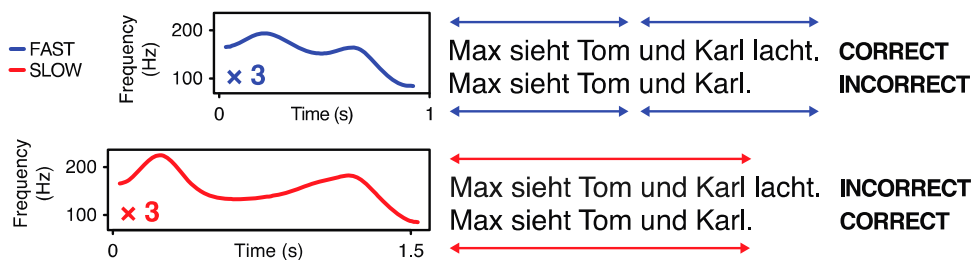


Figure 2. Paradigm (Experiment 1). Prosodic contours repeated 3 times to induce entrainment, followed by time-matched rapid serial visual presentation (RSVP) of target sentence. FAST entrainment aims at inducing correct segmentation of LONG sentence, but incorrect segmentation of SHORT sentence. SLOW entrainment, on the contrary, aims at inducing correct segmentation of SHORT sentence, but incorrect segmentation of LONG sentence.

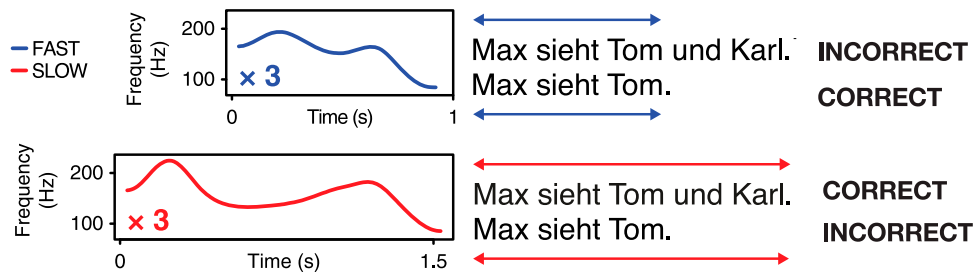


Figure 3. Paradigm (Experiment 2). Prosodic contour repeated 3 times to induce entrainment, followed by time-matched RSVP of the target sentence. According to the prediction hypothesis, FAST entrainment is thought to facilitate the comprehension of SHORT sentences, but inhibit the comprehension of LONG sentences. SLOW entrainment, on the contrary, is expected to facilitate comprehension of LONG sentences, but hinder comprehension of SHORT sentences.

Sentences were modified from Experiment 1. We added surnames to the names to increase sentence duration. We used 153 monosyllabic German surnames matched for length to 3–7 characters; word frequency was normally distributed. Similar-sounding surnames (e.g. *Wolff* and *Wolf*) were not included. On the basis of this list, we constructed surname triplets, none of which contained repeating items, and added one triplet to each of the 6,000 sentences from Experiment 1. Every sentence was combined with a different surname triplet. Adjacent triplets with repeating surnames were avoided. List generation followed the previous experiments.

In LONG sentences (e.g. *Max Scholz sees Tom Schmidt and Karl Weiss laughs*), the last three words (third noun (N3), third surname (SUR3), and second verb (V2); e.g. *Karl, Weiss, and laughs*) were self-paced. In SHORT sentences (e.g. *Max Scholz sees Tom Schmidt and Karl Weiss*), the last two words (N3 and SUR3; e.g. *Karl and Weiss*) were self-paced. The duration of the preceding prosodic contour and thus the frequency of the prior prosodic entrainment was adjusted to the duration of the critical sentence segments: The SLOW contour matched the duration of *Max Scholz sees Tom Schmidt and Karl Weiss*; the FAST contour matched the duration of *Max Scholz sees Tom Schmidt* (Figure 4). Contour generation was identical to Experiments 1 and 2. The duration of the SLOW contour, based on the average

duration of the respective recordings was 1,900 ms; the duration of the FAST contour was 3,040 ms. Average word duration for visual presentation calculated from the synthesised contours length was 320 ms.

Experiment 4

Our fourth experiment tested the entrainment hypothesis on even longer sentences (see Results Experiment 3). Sentences were modified from Experiment 3. We added prepositional phrases (e.g. “*in the city*”) after the verb to increase sentence duration. Every phrase consisted of three words: preposition, article, and noun, all of which were 1-syllable and 3–8 letters in length. Frequency and length of the nouns were normally distributed; the prepositions were 2–3 letters in length (*in, an, auf, vor, bei*) and articles were 3 letters in length (dative case – *der* or *dem*). Each phrase was semantically matched to the corresponding verb pair. List generation matched the previous experiments.

In LONG sentences (e.g. *Max Scholz sees in the city Tom Schmidt and Karl Weiss laughs*), the last three words (third noun (N3), third surname (SUR3), and second verb (V2); e.g. *Karl, Weiss, and laughs*) were self-paced. In SHORT sentences (e.g. *Max Scholz sees in the city Tom Schmidt and Karl Weiss*), the last two words (N3 and SUR3; e.g. *Karl and Weiss*) were self-paced. The duration of the preceding prosodic contour and thus

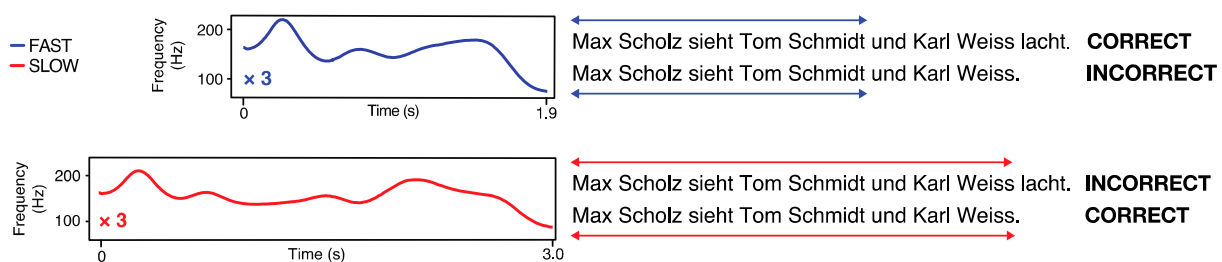


Figure 4. Paradigm (Experiment 3). Prosodic contour repeated 3 times to induce entrainment, followed by time-matched RSVP presentation of the target sentence. FAST entrainment aims at inducing correct segmentation of LONG sentence, but incorrect segmentation of SHORT sentence. SLOW entrainment, on the contrary, aims at inducing correct segmentation of SHORT sentence, but incorrect segmentation of LONG sentence.

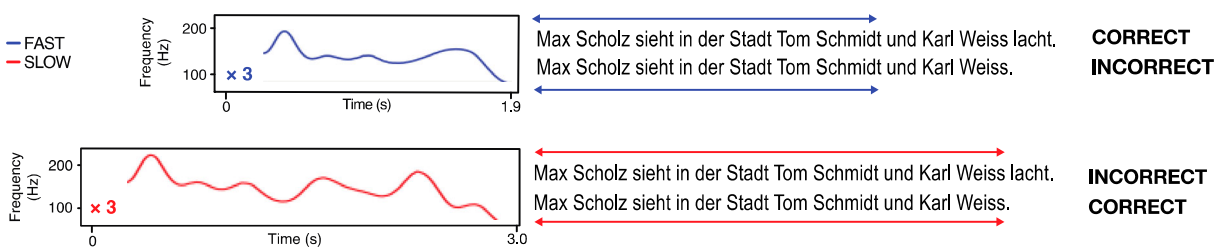


Figure 5. Paradigm (Experiment 4). Prosodic contour repeated 3 times to induce entrainment, followed by time-matched RSVP presentation of the target sentence. FAST entrainment aims at inducing correct segmentation of LONG sentence, but incorrect segmentation of SHORT sentence. SLOW entrainment, on the contrary, aims at inducing correct segmentation of SHORT sentence, but incorrect segmentation of LONG sentence.

the frequency of the prior prosodic entrainment was adjusted to the duration of the critical sentence segments: The SLOW contour matched the duration of *Max Scholz sees in the city Tom Schmidt and Karl Weiss*; the FAST contour matched the duration of *Max Scholz sees in the city Tom Schmidt* (Figure 5). Contour generation matched Experiments 1–3. The duration of the SLOW contour, based on the average duration of the respective recordings was 2,600 ms; the duration of the FAST contour was 3,450 ms. Average word duration for visual presentation calculated from the synthesised contours length was 380 ms.

Procedure

The procedure was identical across the four experiments. Experiments were run online utilising the Gorilla Experiment Builder (www.gorilla.sc; Anwyl-Irvine et al., 2020). Each trial started with a visual fixation cross and auditory presentation of one of the two prosodic contours. Each contour was repeated three times with pauses of 160 ms between repetitions (Ghitza, 2017). For the duration of auditory stimulation in every experiment, see the respective Stimuli section. A pause corresponding to the difference of duration between the FAST and SLOW contours was added before the FAST contour to equalise entrainment phase duration across conditions.

Following a contour, a target sentence (either SHORT or LONG) was presented word by word. The first several words were presented in rapid serial visual presentation (RSVP; Young, 1984). Critically, presentation rate, albeit identical across conditions, was adjusted to contour duration to elicit prosodic transfer effects. Presentation of the final words was self-paced in order to measure those transfer effects (see also Stimuli).

After sentence presentation and a delay of 500 ms, comprehension questions were presented in 75% of trials. To avoid strategy build-up, questions requiring a *yes* answer and questions requiring a *no* answer were both included for each condition. Three kinds of

questions were used: Type 1 – negative for LONG and positive for SHORT (e.g. Exp. 1: *Did Max see Karl?*), Type 2 – positive for both LONG and SHORT (e.g. Exp.1: *Did Max see Tom?*), Type 3 – negative for both LONG and SHORT (e.g. Exp.1: *Did Tom see Max?*). To balance the proportion of expected *yes* and *no* answers, Type 1 and Type 2 were each used in 50% of questions for the LONG sentences. For SHORT, we used Type 1 in 50% of cases and Type 3 in the other 50%.

Participants were instructed to listen to the audio contour, read the sentences presented word-by-word, press a button to continue for the last word(s) of the sentence, and then answer a comprehension question in 75% of trials. There was a response timeout of 2,000 ms. In case of timeout, a screen stating “*Please answer faster*” appeared, and the experiment advanced to the next trial. Transfer effect was inferred from self-paced reading (SPR) times, as well as from reaction times (RTs) and accuracy on comprehension questions.

Data analysis (experiments 1–4)

Data analysis was performed in R (R Core Team, 2019). For the self-paced reading data, in each experiment we focused on 2 critical words: The first word was the last one in a potential segment hypothetically induced by the previous contour (N3 for Experiment 1,3,4 and N2 for Experiment 2), and the second word was the final word of the sentence, where the hypothetical garden-path effect could be discovered (V2 for Experiment 1,3,4 and N3 for Experiment 2).

Outliers were removed separately within the two critical self-paced words and within the comprehension questions, across conditions, according to the interquartile range (outer fences, using 1.5 as a multiplier; Tukey, 1977). Across experiments 1–4, rejection rates for self-paced reading were 10%, 9.48%, 10.85%, and 11.3% respectively; rejection rates for comprehension questions were 0%, 2.16%, 0.02%, and 0.11% respectively. RTs were normalised utilising the Box–Cox

method (Box & Cox, 1964). The effect of entrainment on sentence processing was modelled using a series of mixed-effects regression models, implemented in the lme4 package (Bates et al., 2015). Separate models were built for self-paced reading data at the two critical words as well as for comprehension questions.

For SPR, the predictor was entrainment SPEED (SLOW versus FAST), while the dependent variable was RT. Note that sentence TYPE (LONG versus SHORT) was not included as a factor in the SPR analysis: At the first critical word, the sentences were identical for LONG and SHORT, while second critical word was only present in LONG sentences. For the comprehension questions, predictors were entrainment SPEED, sentence TYPE, and their interaction. Separate model comparisons were performed for RT and accuracy.

Predictors were coded using mean-centered effects coding. Random intercepts were included for subjects and verb pairs. The models with the maximal random-effects structure were attempted initially, but failed to converge in all cases. We therefore used a forward best-path method to determine which random slopes to include, based on established inclusion criteria ($\alpha = 0.2$; Barr et al., 2011). The best-fitting models included by-subject and by-pair-number adjustments to the slopes of entrainment duration and sentence type in specific combinations. The fixed-effects structure and the model comparison path were determined via the predictor structure and the results of initial model comparison. In case of a single predictor (i.e. SPR), we compared the model that included both the predictor as a fixed effect and random effects to the random-effects-only null model. In case of two predictors (i.e. comprehension questions), we first compared single-predictor models separately to the null model. In case any of the comparisons were found significant, we then proceeded to comparing the single-predictor model to the model which included the main effects of both predictors, to see whether the second predictor explain additional variance. In case the combination of both main effects improved model fit, we compared this model against a final model with their interaction included.

Results (experiments 1–4)

Experiment 1

We hypothesised that SLOW entrainment should accelerate SPR at N3 (third noun) – that is, button presses would be accelerated when this word appears on the screen – as it would fall into a single segment with the preceding words (e.g. *[Max sees Tom and Karl]*). Conversely, SPR should decelerate under SLOW entrainment at

V2 (second verb), which would be unexpected (e.g. *[Max sees Tom and Karl] laughs*). For comprehension questions, we expected higher RTs for LONG sentences, as they would produce a garden-path effect, and for this effect to be further strengthened under SLOW entrainment, which would reinforce the incorrect segmentation.

At N3, SPEED led to a significant model improvement: On average, participants were 8.32 ms faster under SLOW as compared to FAST entrainment ($\chi^2(1) = 7.07$, $p = 0.008$; Figure 6a left). No effects were found at V2 ($\chi^2(1) = 1.88$, $p = 0.170$, Figure 6a right). In RTs to comprehension questions, TYPE led to model improvement: SHORT sentences were generally processed slower than LONG sentences ($\chi^2(1) = 260.7$, $p < 0.001$), indicating increased processing difficulty (Figure 6b left). SPEED did not improve model fit ($\chi^2(1) = 0.32$, $p = 0.572$), and neither did the TYPE \times SPEED interaction ($\chi^2(1) = 1.58$, $p = 0.208$). For accuracy, TYPE also caused model improvement: participants were generally more accurate on SHORT sentences than LONG sentences ($\chi^2(1) = 19.03$, $p < 0.001$; Figure 6b right). Note that this pattern appears to indicate a speed–accuracy trade-off effect: the longer the RT, the higher was the accuracy, and vice versa (Davidson & Martin, 2013; Martin & McElree, 2018; Van Dyke & McElree, 2011). This led us to conclude that trials with higher RT had a higher cognitive load and required more resources for achieving a desired accuracy level. Moreover, neither SPEED ($\chi^2(1) = 1.92$, $p = 0.166$) nor the TYPE \times SPEED interaction ($\chi^2(1) = 1.02$, $p = 0.312$) improved model fit for accuracy.

While these results show an influence of prosodic entrainment on processing, they do not support a direct influence on segmentation as such: If the sentence had been expected to end after N3, we should have obtained increased SPR times at V2 due to unexpectedness.

One possible alternative interpretation is that entrainment triggers an approximate temporal prediction of the duration of the upcoming segment or sentence. This converges on an earlier report that found prosodic context to enable the prediction of the count of remaining upcoming words within a sentence (Grosjean, 1983). This study presented subjects with sentences cut off at different timepoints of a target word. Participants could determine the count of the words of the entire sentence based on preceding prosodic cue. Critically, count predictions were only approximate, covering a range of multiple counts around the correct one. The “duration prediction” account therefore claims that subjects respond faster to stimuli at the end of approximately predicted temporal intervals, irrespective of the sentence content and segmentation. Furthermore, this effect is not exclusive to sentence comprehension: it

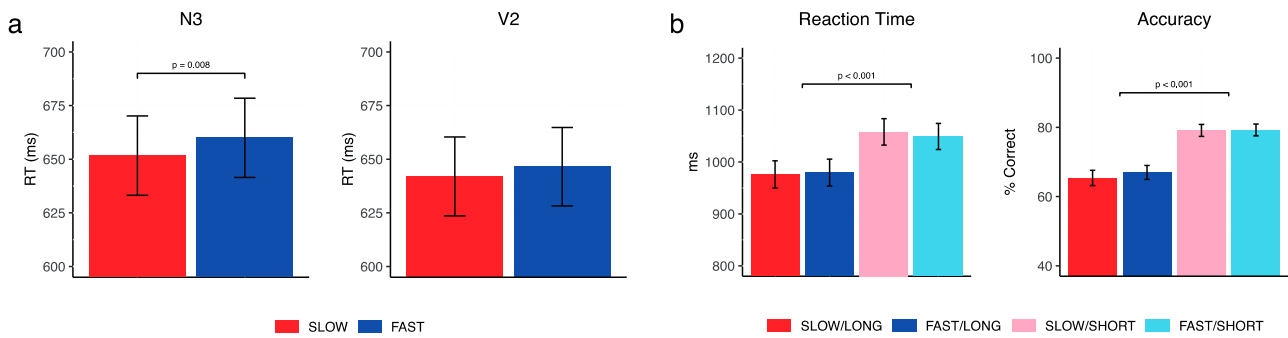


Figure 6. Results (Experiment 1). (a) Self-paced reading: speed-up at N3 under SLOW entrainment, no significant effect of entrainment at V2. (b) Comprehension questions: a main effect of TYPE.

has also been validated for low-level stimuli, like acoustic tones (Stefanics et al., 2010) and recently further replicated (Herbst et al., 2022b). Moreover, the same studies have shown that the reaction times in response to these stimuli correlate with the phase of the delta band oscillation in the brain. Importantly in the context of current study, the prediction mechanism is not always precise in terms of segments; it constitutes a rough bet, whether the upcoming stimuli will be shorter or longer (cf. Grosjean, 1983), unlike the segmentation mechanism, which is precise in the length of the upcoming constituents. In the context of our study, participants may have defaulted to always expecting a LONG sentence (i.e. 50% of sentences continued after N3), as evident from their lower RTs on comprehension questions to LONG sentences. SLOW entrainment may have strengthened this default. Under FAST entrainment, however, the prediction for a longer sentence is weakened; this is demonstrated by higher RTs at N3.

Our pattern could also be captured by a second alternative interpretation, under which the rhythmic prosodic context directs the attentional focus towards a specific future time point (Calderone et al., 2014; Lakatos et al., 2007, 2019). This “temporal attention” account also presupposes temporal prediction, which is, however, not directed to segment duration, but

rather specific points in time. In our case, this account could capture the pattern of an effect at the N3 time point in spite of no detectable garden-path effect at the time point of V2.

To test the temporal attention hypothesis further, we conducted Experiment 2, which employed short non-ambiguous sentences without involving a hypothesis on segmentation. Additionally, this second experiment helped investigating whether the effect of prosodic entrainment is confined to a certain range of sentence durations.

Experiment 2

Under the temporal attention account, for the SPR, we expected a speed-up at N2 (second noun) for the FAST contour, similar to the N3 (third noun) effect in Experiment 1. In addition, we hypothesised a speed-up at N3 (third noun) for SLOW contours in LONG sentences. For comprehension questions, we expected lower RTs for LONG sentences after SLOW contours and higher RTs after FAST contours, since the SLOW contour would draw attentional focus to the end of the LONG sentence, facilitating its processing. On the contrary, for the SHORT sentences we expected lower RTs under the FAST entrainment and higher RTs under SLOW, since the attentional

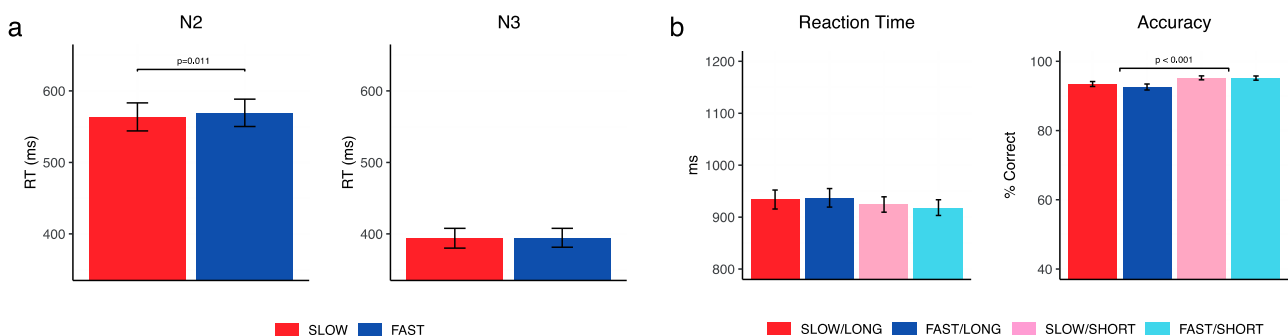


Figure 7. Results (Experiment 2). (a) Self-paced reading: speed-up at N3 (3rd noun) under SLOW entrainment, no significant effect of entrainment at V2 (2nd verb). (b) Comprehension questions; main effect of TYPE for accuracy.

focus of FAST contour would be precisely at the end of a SHORT sentence. We expected the accuracy data to follow the same pattern as RTs, due to the speed–accuracy trade-off we found in Experiment 1.

Contrary to the SPR hypotheses, RT at N2 was 5.65 ms faster under SLOW as compared to FAST entrainment; SPEED improved model fit ($\chi^2(1)=6.52$, $p=0.011$; Figure 7a left). At N3, SPEED did not improve model fit ($\chi^2(1)=0.48$, $p=0.487$, Figure 7a right). For RTs to comprehension questions (Figure 7b left), models did not improve under inclusion of either TYPE ($\chi^2(1)=1.60$, $p=0.204$), SPEED ($\chi^2(1)=0.19$, $p=0.666$), or the TYPE \times SPEED interaction ($\chi^2(1)=1.69$, $p=0.193$). In accuracy data (Figure 7b right), TYPE led to model improvement: participants were generally more accurate on SHORT sentences than LONG sentences ($\chi^2(1)=13.22$, $p<0.001$), similar to Experiment 1. There was no improvement of accuracy models under inclusion of either SPEED ($\chi^2(1)=1.97$, $p=0.160$) or the TYPE \times SPEED interaction ($\chi^2(1)=0.85$, $p=0.358$).

Again, the results confirm an influence of the rhythmic prior prosodic stimulus on sentence processing. The specific time point of the effect within the sentence, however, argues against an explanation in terms of temporal attention. Rather, it appears to support the duration prediction account: The N2 speedup for the SLOW condition appears to indicate that subjects expected the sentence to continue into a LONG sentence, when they had heard a SLOW prosodic contour before. In contrast, the N2 effect for the SLOW contour falsifies the temporal attention account, which did predict this effect for FAST instead. Moreover, the temporal attention account does not predict a speedup for SLOW contour at N2 – instead, it would have required a speedup at N3. As to the isolated main effect of TYPE for accuracy without an RT decrease, it might be an artefact of the experimental design: as the SHORT sentences consisted of only three words in Experiment 2 (e.g. *Max sees Tom*), they were very easy to process for the participants.

The numeric decrease of the SPR effect relative to Experiment 1 may indicate that entrainment depends on sentence duration. Alongside testing our working hypothesis of a duration prediction, we thus reconsider our original segmentation hypothesis in Experiment 3: Insufficient sentence duration in Experiments 1 and 2 could have decreased potential garden-path effects at the verb. Consequently, we employed longer versions of the sentences from Experiment 1 again. Should we observe an SPR effect at the verb, we would conclude that prosodic entrainment affects segmentation only at a certain minimum sentence duration.

Experiment 3

Should the segmentation account be valid, FAST entrainment would induce correct segmentation of LONG sentences (= lower RTs), but incorrect in case of SHORT sentences (= higher RTs). SLOW entrainment, in turn, would induce correct segmentation of SHORT sentences (lower RTs), but incorrect in case of LONG sentences (= higher RTs). On the contrary, should the duration prediction account be correct, the processing of the LONG sentences would be facilitated by SLOW contour (= lower RTs) and inhibited by FAST (= higher RTs), while the processing of the SHORT sentences would be facilitated by the FAST contour (= lower RTs) and hindered by the SLOW contour (= higher RTs). We also expected the accuracy to be proportional to RTs according to the speed–accuracy trade-off effect found in Experiment 1, no matter the supported account. Regarding the self-paced reading data, both accounts would predict a speed-up effect at N3 (third noun) for the SLOW contour. However, only the segmentation account generates a strong hypothesis for a slow-down at V2 (second verb) for the SLOW contour.

The speed-ups at N3 (Experiment 1) and N2 (Experiment 2) were replicated at N3 in Experiment 3: On average, participants were 18.61 ms faster under

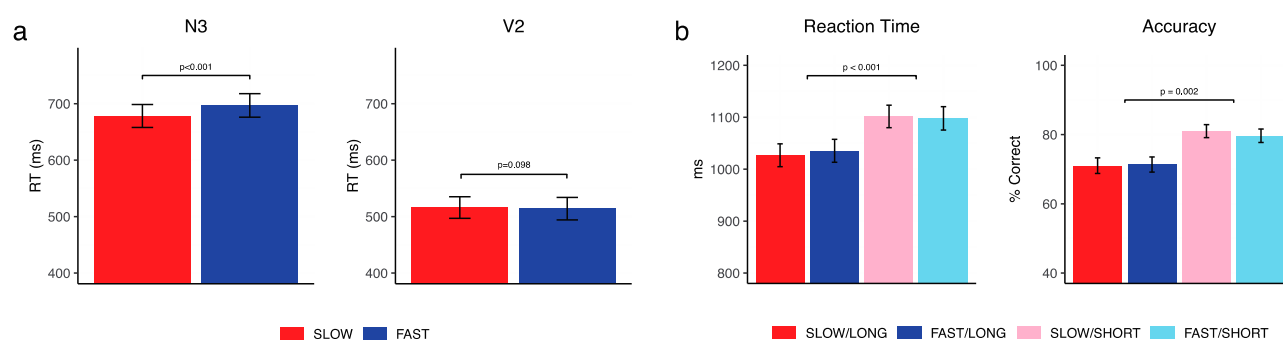


Figure 8. Results (Experiment 3). (a) Self-paced reading: speed-up at N3 (3rd noun) under SLOW entrainment, combined with a trending garden-path effect, strengthened by entrainment, at V2 (2nd verb). (b) Comprehension questions; main effect of TYPE.

SLOW as compared to FAST entrainment – SPEED improved the model fit ($\chi^2(1) = 18.2$, $p < 0.001$, Figure 8a left). At V2 in the third experiment, SPEED yielded a trending model improvement ($\chi^2(1) = 2.73$, $p = 0.098$, Figure 8a right). An analysis across experiments 1–3 showed that the N3 effect depended on sentence duration (i.e. interaction EXPERIMENT \times SPEED; $\chi^2(1) = 6.35$, $p = 0.042$). Interestingly, an analysis at V2 across experiments 1 and 3 – where V2 was actually present – yielded that EXPERIMENT \times SPEED did improve model fit as well: the longer the sentences, the longer the slowdown under SLOW entrainment ($\chi^2(1) = 4.19$, $p = 0.040$).

In comprehension questions, TYPE improved model fit for RTs ($\chi^2(1) = 36.4$, $p < 0.001$): SHORT sentences were processed slower than LONG sentences (Figure 8b left). Neither the factor of SPEED ($\chi^2(1) = 0.02$, $p = 0.86$) nor the TYPE \times SPEED interaction ($\chi^2(1) = 2.70$, $p = 0.100$) improved the model fit for question RTs. For accuracy (Figure 8b right), we found a significant improvement under the inclusion of TYPE ($\chi^2(1) = 9.71$, $p = 0.002$) – participants were more accurate at SHORT sentences as compared to LONG, replicating the speed-accuracy trade-off found in Experiment 1. No significant effects were found for the inclusion of either SPEED ($\chi^2(1) = 0.74$, $p = 0.388$) or TYPE \times SPEED interaction ($\chi^2(1) = 3.17$, $p = 0.075$).

The speed-up at N3 is consistent with a strengthened prediction for a LONG sentence under SLOW entrainment. Still, this is also consistent with the segmentation hypothesis, since under SLOW entrainment N3 would be putatively in the same segment as the previous words, which could reduce processing speed as well. Across-experiment comparisons suggest that speed-up increases with sentence duration. The V2 trend, together with the increase of the V2 difference across experiments 1 and 3, does provide some limited evidence for the segmentation account, indicating a possible garden-path effect. To test the hypothesis that the

garden-path effect emerges only for longer sentences, we conducted Experiment 4 with increased sentence length.

Experiment 4

In Experiment 4, should the segmentation account be valid, FAST entrainment would induce correct segmentation of LONG sentences (= lower RTs), but incorrect segmentation in case of SHORT sentences (= higher RTs). SLOW entrainment, in turn, would induce correct segmentation of SHORT sentences (= lower RTs), but incorrect in case of LONG sentences (= higher RTs). On the contrary, should the duration prediction account be correct, the processing of the LONG sentences would be facilitated by SLOW contour (= lower RTs) and inhibited by FAST (= higher RTs), while the processing of the SHORT sentences would be facilitated by the FAST contour (= lower RTs) and hindered by the SLOW contour (= higher RTs). We also expected the accuracy to be proportional to RTs according to the speed-accuracy trade-off effect found in Experiment 1, no matter the supported account. Regarding the self-paced reading data, both accounts would predict a speed-up effect at N3 (third noun) for the SLOW contour. However, only the segmentation account would predict a slow-down at V2 (second verb) for the SLOW contour.

For self-paced reading, the speed-ups at N3 (Experiment 1 and 3) and N2 (Experiment 2) were replicated at N3 in Experiment 4: On average, participants were 14.57 ms faster under SLOW as compared to FAST entrainment – SPEED improved the model fit ($\chi^2(1) = 15.85$, $p < 0.001$, Figure 9a left). No effects were found at V2 ($\chi^2(1) = 0.52$, $p = 0.471$, Figure 9a right); numerically, there was not a delay, but a speed-up of 0.4 ms for the SLOW entrainment, contrary to the predictions of the segmentation account. The results of the Experiment 4 disrupted the pattern, observed in the previous three experiments: average speed-up/delay in the RTs is

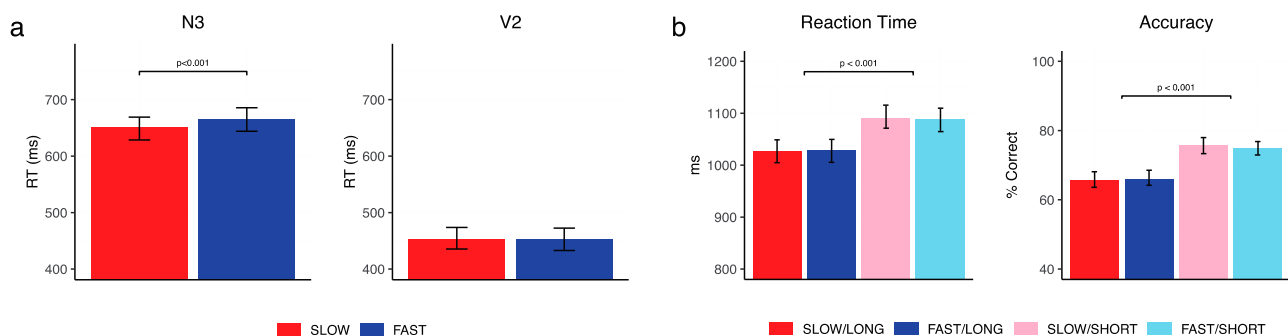


Figure 9. Results (Experiment 4). (a) Self-paced reading: speed-up at N3 (3rd noun) under SLOW entrainment, no significant effect of entrainment at V2 (2nd verb). (b) Comprehension questions; main effect of TYPE.

shorter for both N3 and V2 in Experiment 4 (14.57 ms and 0.4 ms, respectively) than in Experiment 3 (18.61 and 1.9 ms, respectively).

In comprehension questions, TYPE improved model fit for RTs ($\chi^2(1) = 199.18$, $p < 0.001$): SHORT sentences were processed more slowly than LONG sentences (Figure 9b left). Neither the factor of SPEED ($\chi^2(1) = 0.006$, $p = 0.94$) nor the TYPE \times SPEED interaction ($\chi^2(1) = 0.13$, $p = 0.71$) improved the model fit for question RTs. For accuracy (Figure 9b right), we also found a significant improvement under the inclusion of TYPE ($\chi^2(1) = 10.97$, $p < 0.001$) – participants were more accurate at SHORT sentences as compared to LONG, replicating the speed–accuracy trade-off found in previous experiments. No significant effects were found for the inclusion of SPEED ($\chi^2(1) = 1.58$, $p = 0.209$). We found, however, model improvement under the inclusion of the TYPE \times SPEED interaction ($\chi^2(1) = 5.43$, $p = 0.0198$). Taken together with the above-mentioned speed–accuracy trade-off, this supports the duration prediction account: It was harder for participants to process LONG sentences under FAST entrainment, and SHORT sentences under SLOW entrainment (Figure 8b right).

The replicated N3 speed-up can be explained by both the duration prediction and segmentation accounts. However, the segmentation account would predict a slowdown at V2, which we failed to observe in Experiments 1 and 4. While these null effects do not allow us to draw a definitive conclusion, their combination with the time point of the SPR effect in Experiment 2 rather speaks in favour of a tentative interpretation in terms of duration prediction. Critically, since the results of Experiment 4 do not fall in line with the trend observed in previous experiments (increase of the speed-up/delay in RTs together with the sentence length), we can no longer claim that the prediction effect is strengthened or a garden-path effect is emerging as the duration of the sentence increases.

Nevertheless, we should note that another type of pattern emerges across the four experiments. As shown in Table 2, the duration prediction effect appears to be strongest at the prosodic contour duration of around 2.5 s. The 2.5–2.7 s limit has been previously associated with other important restraints in psycholinguistics, such as median sentence duration in speech

(Vollrath et al., 1992) and a general time window for perception, simultaneity, or information integration (Baddeley et al., 1975; Pöppel, 1997; White, 2017). A possible explanation of this preference could be linked to working memory constraints; yet another interpretation could be connected to constraints on electrophysiological level. In particular, entrainment could potentially only be possible within a particular electrophysiological frequency band. Further electrophysiological research is needed to investigate this hypothesis.

General discussion

We investigated here whether entrainment to prosody influences the processing of a subsequent sentence. We initially hypothesised that entrainment would facilitate the segmentation of upcoming sentences, in line with the function of prosody in bottom-up segmentation. While the results suggest that prosodic entrainment does affect the processing of subsequent sentences, the specific pattern speaks against this hypothesis. The results rather suggest that that prosodic entrainment allows for a prediction of the duration of the upcoming sentence; potentially, it even allows to predict the duration of syntactic segments.

In general, a duration prediction interpretation converges on earlier psycholinguistic conceptualisations (Breen, 2014; Frazier et al., 2006; Grosjean, 1983): According to the rational speaker hypothesis, listeners interpret strong IPBs as predictive of longer upcoming segments (Carlson et al., 2001; Watson & Gibson, 2004). While our behavioural results do not allow for strong links to possible neuronal underpinnings of such effects, they are consistent with the finding that oscillations in the delta range entrain to rhythmic speech stimuli, triggering the estimation of the duration of subsequent time intervals between stimulation offset and the occurrence of upcoming auditory targets (Breska & Deouell, 2017; Herbst et al., 2022; Stefanics et al., 2010). Future work using electrophysiological methodology is required to ascertain that the repetitive prosodic carrier wave in the current study elicited a persisting oscillation that continued after the offset of the auditory stimulus, facilitating the comprehension of subsequent visual sentences through a prediction of their duration.

As already mentioned in the Introduction, we distinguish the current usage of the term *entrainment* from its usage to denote the general adjustment of two speakers to each other in a dialogue. From a broader perspective, however, the results of our study are also relevant in face-to-face communication between speakers. As known from previous work, prosodic context is rhythmic (Stehwien & Meyer, 2022) and

Table 2. Reaction time effect pattern across four experiments.

Experiment	Contour duration (s)	RT speed-up in SLOW condition (ms)
1–2	1.5	7 (averaged)
3	2.5	19
4	3.5	15

can therefore facilitate speech processing by providing tentative predictions of upcoming sentence/segment length (Fodor, 2002; Frazier et al., 2006; Steinhauer & Friederici, 2001). It has also been previously shown on syllable level that rhythm in the context may constitute a tool for the listener to facilitate processing in a dialogue, providing conversational entrainment (Wilson & Wilson, 2005). As to neural entrainment, it is known to operate across levels, such as phonetic, syntactic, prosodic, and finally, discourse- and referential-level processing (Meyer et al., 2019, 2020). Moreover, latest evidence shows that conversational entrainment operates on prosodic level as well (Levitan et al., n.d.; Levitan & Hirschberg, n.d.; Reichel et al., 2018). Therefore, speech rhythms could facilitate conversational entrainment via neural entrainment: speakers may use rhythmic prosodic cues to predict the duration of the interlocutor's upcoming phrases, thus improving and speeding up the communication process. Future studies could address the question whether speakers do also actively manage prosodic cues, as well as the possible neural underpinnings of this effect. Another interesting research direction could look into which particular speech features that speakers align during conversation are periodic enough to be exploited by periodic and possibly predictive processes.

The current results do not provide strong support for our initial segmentation hypothesis: If entrainment affected segmentation, sentence-final verbs in LONG sentences should have been harder to process under SLOW entrainment. This so-called garden-path effect has been observed previously for the coordination ambiguities used in the present study (Henke & Meyer, 2020; Hoeks, Hendriks, Vonk, Brown, & Hagoort, 2006; Hoeks et al., 2002). In Experiment 3, we obtained a trend for a modulation of this effect via entrainment; however, that was subsequently refuted by Experiment 4, where even longer sentences did not show such an effect. Our results also do not support an explanation in terms of priming (see Introduction), as in the event of direct priming of the sentence length by the prosodic contour, we would have had a garden-path effect for the second verb in the LONG sentences under the SLOW prosodic condition, which was not the case for any of the experiments in the series.

Acknowledgements

The authors would like to thank Lena Henke, Katie McCann, Katharina Menn, Laura Riedel and Antonia Schmidt for the assistance in stimuli preparation and pilot data recording, as well as Lorenzo Titone for the assistance in editing the manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The Max Planck Society funded this research.

Data availability

All data, analysis code, and research materials are available on request.

ORCID

Yulia Lamekina  <http://orcid.org/0000-0002-7016-038X>

References

- Aarts, A. A., Anderson, J. E., Anderson, C. J., Attridge, P. R., Attwood, A., Axt, J., Babel, M., Bahník, Š, Baranski, E., Barnett-Cowan, M., Bartmess, E., Beer, J., Bell, R., Bentley, H., Beyan, L., Binion, G., Borsboom, D., Bosch, A., Bosco, F. A., ... Zuni, K. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716–aac4716. <https://doi.org/10.1126/science.aac4716>
- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52(1), 388–407. <https://doi.org/10.3758/s13428-019-01237-x>
- Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 14(6), 575–589. [https://doi.org/10.1016/S0022-5371\(75\)80045-4](https://doi.org/10.1016/S0022-5371(75)80045-4)
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2011). Random effects structure in mixed-effects models: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Boersma, P., & van Heuven, V. (2001). Speak and unSpeak with Praat. *Glott International*, 5(9–10), 341–347.
- Bosker, H. R. (2017). Accounting for rate-dependent category boundary shifts in speech perception. *Attention, Perception, and Psychophysics*, 79(1), 333–343. <https://doi.org/10.3758/s13414-016-1206-4>
- Bourguignon, M., De Tiège, X., De Beeck, M. O., Ligot, N., Paquier, P., Van Bogaert, P., Goldman, S., Hari, R., & Jousmäki, V. (2013). The pace of prosodic phrasing couples the listener's cortex to the reader's voice. *Human Brain Mapping*, 34(2), 314–326. <https://doi.org/10.1002/hbm.21442>
- Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2), 211–243. <http://dx.doi.org/10.1111/rssb.1964.26.issue-2>

- Branigan, H. P. (2007). Syntactic priming. *Language and Linguistics Compass*, 1(1–2), 1–16. <https://doi.org/10.1111/j.1749-818x.2006.00001.x>
- Breen, M. (2014). Empirical investigations of the role of implicit prosody in sentence processing. *Linguistics and Language Compass*, 8(2), 37–50. <https://doi.org/10.1111/lnc3.12061>
- Breen, M., Kaswer, L., Van Dyke, J. A., Krivokapić, J., & Landi, N. (2016). Imitated prosodic fluency predicts reading comprehension ability in good and poor high school readers. *Frontiers in Psychology*, 7(JUL), 1026. <https://doi.org/10.3389/fpsyg.2016.01026>
- Brennan, S. E. (1996). Lexical entrainment in spontaneous dialog. *Proceedings, 1996 International Symposium on Spoken Dialogue, ISSD-96*, 41–44.
- Breska, A., & Deouell, L. Y. (2017). Neural mechanisms of rhythm-based temporal prediction: Delta phase-locking reflects temporal predictability but not rhythmic entrainment. *PLoS Biology*, 15(2), 1–30. <https://doi.org/10.1371/journal.pbio.2001665>
- Brown, M., Salverda, A. P., Dille, L. C., & Tanenhaus, M. K. (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychonomic Bulletin and Review*, 18(6), 1189–1196. <https://doi.org/10.3758/s13423-011-0167-9>
- Calderone, Daniel J, Lakatos, Peter, Butler, Pamela D., & Castellanos, F. Xavier. (2014). Entrainment of neural oscillations as a modifiable substrate of attention. *Trends in Cognitive Sciences*, 18(6), 300–309. <http://dx.doi.org/10.1016/j.tics.2014.02.005>
- Carlson, K., Clifton, C., & Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language*, 45(1), 58–81. <https://doi.org/10.1006/jmla.2000.2762>
- Clifton, C., Carlson, K., & Frazier, L. (2002). Informative prosodic boundaries. *Language and Speech*, 45(2), 87–114. <https://doi.org/10.1177/00238309020450020101>
- Davidson, D.J., & Martin, A.E. (2013). Modeling accuracy as a function of response time with the generalized linear mixed effects model. *Acta Psychologica*, 144(1), 83–96. <http://dx.doi.org/10.1016/j.actpsy.2013.04.016>
- Dille, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, 59(3), 294–311. <https://doi.org/10.1016/j.jml.2008.06.006>
- Dille, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science*, 21(11), 1664–1670. <https://doi.org/10.1177/0956797610384743>
- Edlund, J. (2011). In search for the conversational homunculus: serving to understand spoken human face-to-face interaction. <http://www.diva-portal.org/smash/record.jsf?pid=diva2:402997>
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Fodor, J. D. (2002). Psycholinguistics cannot escape prosody. In *Proceedings of the 1st International Conference on Speech Prosody, April 2002*, 83–88.
- Frazier, L., Carlson, K., & Clifton, C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, 10(6), 244–249. <https://doi.org/10.1016/j.tics.2006.04.002>
- Ghitza, O. (2017). Acoustic-driven delta rhythms as prosodic markers. *Language, Cognition and Neuroscience*, 32(5), 545–561. <https://doi.org/10.1080/23273798.2016.1232419>
- Grosjean, F. (1983). How long is the sentence? Prediction and prosody in the on-line processing of language. *Linguistics*, 21(3), 501–530. <https://doi.org/10.1515/ling.1983.21.3.501>
- Heister, J., Würzner, K. M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). dlexDB – Eine lexikalische datenbank für die psychologische und linguistische forschung. *Psychologische Rundschau*, 62(1), 10–20. <https://doi.org/10.1026/0033-3042/a000029>
- Henke, L., & Meyer, L. (2020). Endogenous Oscillations Time-Constrain Linguistic Segmentation: Cycling the Garden Path. *Cerebral Cortex (in press)*.
- Henry, M. J., & Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proceedings of the National Academy of Sciences*, 109(49), 20095–20100. <https://doi.org/10.1073/pnas.1213390109>
- Herbst, S. K., Stefanics, G., & Obleser, J. (2022). Endogenous modulation of delta phase by expectation – A replication of stefanics et al., 2010. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior*, 149, 226–245. <https://doi.org/10.1016/J.CORTEX.2022.02.001>
- Hickok, G., Farahbod, H., & Saberi, K. (2015). The rhythm of perception: Entrainment to acoustic rhythms induces subsequent perceptual oscillation. *Psychological Science*, 26(7), 1006–1013. <https://doi.org/10.1177/0956797615576533>
- Hoeks, J. C. J., Hendriks, P., Vonk, W., Brown, C. M., & Hagoort, P. (2006). Processing the noun phrase versus sentence coordination ambiguity: Thematic information does not completely eliminate processing difficulty. *Quarterly Journal of Experimental Psychology*, 59(9), 1581–1599. <https://doi.org/10.1080/17470210500268982>
- Hoeks, J. C. J., Vonk, W., & Schriefers, H. (2002). Processing coordinated structures in context: The effect of topic-structure on ambiguity resolution. *Journal of Memory and Language*, 46(1), 99–119. <https://doi.org/10.1006/jmla.2001.2800>
- Inbar, M., Grossman, E., & Landau, A. N. (2020). Sequences of intonation units form a 1 Hz rhythm. *BioRxiv*, 10(1), 1–23. <https://doi.org/10.1101/765016>
- Klasmeyer, G. (1997). Perceptual importance of selected voice quality parameters. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing – Proceedings*, <https://doi.org/10.1109/ICASSP.1997.598808>.
- Kösem, A., Bosker, H. R., Takashima, A., Meyer, A. S., Jensen, O., & Hagoort, P. (2018). Neural entrainment determines the words we hear. *Current Biology*, 28(18), 2867–2875. <https://doi.org/10.1016/j.cub.2018.07.023>
- Kotz, S. A., Meyer, M., Alter, K., Besson, M., Von Cramon, D. Y., & Friederici, A. D. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain and Language*, 86(3), 366–376. [https://doi.org/10.1016/S0093-934X\(02\)00532-1](https://doi.org/10.1016/S0093-934X(02)00532-1)
- Lakatos, Peter, Chen, Chi-Ming, O’Connell, Monica N, Mills, Aimee, & Schroeder, Charles E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron*, 53(2), 279–292. <http://dx.doi.org/10.1016/j.neuron.2006.12.011>
- Lakatos, Peter, Gross, Joachim, & Thut, Gregor. (2019). A new unifying account of the roles of neuronal entrainment.

- Current Biology*, 29(18), R890–R905. <http://dx.doi.org/10.1016/j.cub.2019.07.075>
- Lehnert-LeHouillier, H., Terrazas, S., & Sandoval, S. (2020). Prosodic entrainment in conversations of verbal children and teens on the autism spectrum. *Frontiers in Psychology*, 11, 2718. <https://doi.org/10.3389/fpsyg.2020.582221>
- Levitan, R., Gravano, A., Willson, L., Be, S., Nü, S., Hirschberg, J., & Nenkova, A. (n.d.). *Acoustic-prosodic entrainment and social behavior*. Retrieved from June 14, 2022, from <http://www.mturk.com>
- Levitan, R., & Hirschberg, J. (n.d.). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions.
- Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, 54(6), 1001–1010. <https://doi.org/10.1016/j.neuron.2007.06.004>
- Mai, G., Minett, J. W., & Wang, W. S. Y. (2016). Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing. *NeuroImage*, 133, 516–528. <https://doi.org/10.1016/j.neuroimage.2016.02.064>
- Martin, Andrea E, & McElree, Brian. (2018). Retrieval cues and syntactic ambiguity resolution: speed-accuracy tradeoff evidence. *Language, Cognition and Neuroscience*, 33(6), 769–783. <http://dx.doi.org/10.1080/23273798.2018.1427877>
- Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: State of the art and emerging mechanisms. *European Journal of Neuroscience*, 48(7), 2609–2621. <https://doi.org/10.1111/ejn.13748>
- Meyer, L., Sun, Y., & Martin, A. E. (2019). Synchronous, but not entrained: Exogenous and endogenous cortical rhythms of speech and language processing. *Language, Cognition and Neuroscience*, 0(0), 1–11. <https://doi.org/10.1080/23273798.2019.1693050>
- Meyer, L., Sun, Y., & Martin, A. E. (2020). “Entraining” to speech, generating language? *Language, Cognition and Neuroscience*, 0(0), 1–11. <https://doi.org/10.1080/23273798.2020.1827155>
- Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., & Von Cramon, D. Y. (2002). fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Human Brain Mapping*, 17(2), 73–88. <https://doi.org/10.1002/hbm.10042>
- New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English lexicon project. *Psychonomic Bulletin and Review*, 13(1), 45–52. <https://doi.org/10.3758/BF03193811>
- Oord, A. v. d., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). WaveNet: A generative model for raw audio. <http://arxiv.org/abs/1609.03499>
- Pannekamp, A., Toepel, U., Alter, K., Hahne, A., & Friederici, A. D. (2005). Prosody-driven sentence processing: An event-related brain potential study. *Journal of Cognitive Neuroscience*, 17(3), 407–421. <https://doi.org/10.1162/0898929053279450>
- Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin*, 134(3), 427–459. <https://doi.org/10.1037/0033-2909.134.3.427>
- Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends in Cognitive Sciences*, 1(2), 56–61. [https://doi.org/10.1016/S1364-6613\(97\)01008-5](https://doi.org/10.1016/S1364-6613(97)01008-5)
- R Core Team. (2019). R: A language and environment for statistical computing. Accessed 1st April 2019.
- Reichel, U. D., Mády, K., & Cole, J. (2018). Prosodic entrainment in dialog acts. <https://doi.org/10.48550/arxiv.1810.12646>
- Snedeker, J., & Casserly, E. (2010). Is it all relative? Effects of prosodic boundaries on the comprehension and production of attachment ambiguities. *Language and Cognitive Processes*, 25(7–9), 1234–1264. <https://doi.org/10.1080/01690960903525499>
- Sonntag, G. P., & Portele, T. (1998). PURR – A method for prosody evaluation and investigation. *Computer Speech and Language*, 12(4), 437–451. <https://doi.org/10.1006/csla.1998.0107>
- Stefanics, G., Hangya, B., Hernádi, I., Winkler, I., Lakatos, P., & Ulbert, I. (2010). Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *Journal of Neuroscience*, 30(41), 13578–13585. <https://doi.org/10.1523/JNEUROSCI.0703-10.2010>
- Stehwien, S., & Meyer, L. (2022). Short-term periodicity of prosodic phrasing. *Speech Prosody*. <https://doi.org/10.31234/osf.io/c9sgb>
- Steinhauer, K., & Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research*, 30(3), 267–295. <https://doi.org/10.1023/A:1010443001646>
- Teoh, E. S., Cappelloni, M. S., & Lalor, E. C. (2019). Prosodic pitch processing is represented in delta-band EEG and is dissociable from the cortical tracking of other acoustic and phonetic features. *European Journal of Neuroscience*, 50(11), 3831–3842. <https://doi.org/10.1111/ejn.14510>
- Tukey, J. W.. (1977). *Exploratory Data Analysis* (Vol. 2, pp. 131–160). <https://doi.org/10.1007/978-1-4419-7976-6>
- Van Dyke, Julie A, & McElree, Brian. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, 65(3), 247–263. <http://dx.doi.org/10.1016/j.jml.2011.05.002>
- Vollrath, M., Kazenwadel, J., & Krüger, H.-P. P. (1992). A universal constant in temporal segmentation of human speech. *Naturwissenschaften*, 79(10), 479–480. <https://doi.org/10.1007/bf01139205>
- Watson, D. G., & Gibson, E. A. F. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, 19(6), 713–755. <https://doi.org/10.1080/01690960444000070>
- Weidman, S., Breen, M., & Haydon, K. C. (2016). Prosodic speech entrainment in romantic relationships. *Proceedings of the International Conference on Speech Prosody*, <https://doi.org/10.21437/speechprosody.2016-104>
- White, P. A. (2017). The three-second “subjective present”: A critical review and a new proposal. *Psychological Bulletin*, 143(7), 735–756. <https://doi.org/10.1037/bul0000104>
- Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review*, 12(6), 957–968. <https://doi.org/10.3758/BF03206432>
- Young, S. R. (1984). RSVP: A task, reading aid, and research tool. *Behavior Research Methods, Instruments, & Computers*, 16(2), 121–124. <https://doi.org/10.3758/BF03202369>