Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

Prosodic prominence and its hindering effect on word recognition memory in German

Barbara Zeyer[1], *Martina Penke*[1]

[1]University of Cologne
barbara.zeyer@uni-koeln.de, martina.penke@uni-koeln.de

Abstract

The aim of our study was to deepen our understanding of the influence of prosodic prominence on language processing. We conducted a word recognition memory task in German where we manipulated the word preceding the target word with either the highest (*L+H**) or the lowest prominent accent type (*deaccentuation*) in German. Previous studies have shown higher accuracy rates when the target word was prosodically manipulated with a high prominent accent type. Based on these findings, we postulated that prosodic prominence binds processing resources, as the attention of the listener is drawn to the prosodically prominent entity, furthering recognition memory of this entity. Conversely, we assumed that a highly prominent accent (*L+H**) on the word preceding the target word would lead to less accurate recognition rates of the target word compared to a control condition with deaccentuation. In this case, processing of the prominently accented word would bind attention and processing resources that would not be available for a deeper anchoring of the following word, the target, in memory. Our data confirmed this expectation. Prosodic prominence hindered recognition when it was on the word preceding the target word. This finding supports our assumption that prosodic prominence binds attention and processing resources.

Index Terms: word recognition, memory, language processing, prosodic prominence, German.

1. Introduction

Prosody plays an important role during conversation. For example, prosody is used to indicate turn-taking [1, 2], to mark information status [3, 4], or to distinguish between background and focus [5, 6]. Additionally, prosody can also be used to highlight particular entities within an utterance in order to draw a listeners' attention to this entity. This highlighting is called prosodic prominence and the prosodically prominent entities stand out of their environment based on their prosodic characteristics [7]. One of these prosodic characteristics is for example pitch accent type.

Pitch accent types differ quite substantially in their perceived prominence. [8] asked participants to rate how highlighted, thus prominent, the proper name *Lana* in sentences like *Sie hat mit der Lana telefoniert* ('She phoned with Lana') sounded. Based on the results of their study, [8] established a scale of perceptual prominence in German, where the pitch accent type *L+H** was rated as most prominent and *deaccentuation* was rated as least prominent.

A number of psycholinguistic studies have revealed effects of prosodic prominence on language processing. Studies have, for instance, examined effects of prosodic prominence on memory in word recognition memory studies (see e.g., [9], [10] for recognition in sentences; see e.g., [11] for recall in serial lists). These studies repeatedly showed that prosodically prominent words are recognized and recalled better than words which are not prosodically prominent.

For instance, [9] examined the effect of two different accent types – *L+H** and *H** – on word recognition memory in experimental sentences that were embedded in a short discourse in American English. The authors found that the more prominent accent type *L+H** led to more accurate recognition rates than the less prominent accent type *H**, demonstrating improved memory in the condition *L+H** which the authors attributed to facilitated encoding in discourse [9].

Similar results were reported by [10], who examined the effects of syntactic and prosodic prominence on word recognition memory in Korean and Australian English. For English speakers, the authors found better recognition memory in the condition where the target word was prosodically prominent (i.e., marked by a high pitch accent) compared to the baseline condition. They attributed this effect to facilitated lexical processing, resulting from listeners' paying more attention to the prosodically prominent entity in an utterance.

In sum, the aforementioned studies show that prosodic prominence can facilitate word recognition memory. Based on these findings, it has been argued that listeners pay more attention to prosodically prominent entities, which facilitates certain aspects of language processing and leads to more accurate word recognition rates for a prosodically prominent target word (see e.g., [10]). However, while studies so far have emphasized the facilitating effects of prosody on word recall, it is unclear whether processing prosodically prominent pitch accents may also come at a price. Thus, it is possible that the processing of prosodically prominent words also binds processing resources, thereby leading to a decrease of processing resources for subsequent words. In the present study, we sought to address this open question by examining the effect of prosodic prominence on word recognition memory when the word *preceding* the word for which recognition memory is tested (here called the target word) is prosodically prominent. Thus unlike previous studies that investigated the effect of prosodic prominence on the target word itself, here we manipulated the prosodic prominence of the word preceding the target word. If prosodic prominence indeed binds processing resources, we hypothesized that target words that were preceded by a prosodically prominent word (manipulated with the pitch accent type *L+H**) would be recalled less accurately than target words that followed a word that was not prosodically prominent (i.e., a word that was deaccented). To investigate this issue, we conducted a word recognition memory study with German participants.

10.21437/SpeechProsody.2024-229

2. Methodology

2.1. Research questions

We adopted a typical setting for a word recognition memory task, in which participants first listened to a block of sentences and subsequently had to decide whether words presented in isolation had been part of one of the previously presented sentences or not (recognition phase). We wanted to examine if prosodic prominence hinders word recognition memory if manipulated on the word preceding the word for which recognition memory is tested (here the target word). Following the scale of perceptual prosodic prominence established by [8], we used the highly prominent accent type *L+H** as the prosodically prominent condition and the low prominent *deaccentuation* as baseline condition. If the processing of prosodic prominence binds processing resources which are then lacking for the processing of subsequent material, we would expect lower recognition rates in the condition where the highly prominent *L+H** accent is placed on the word preceding the target word for which memory recognition is tested, compared to the baseline condition where the word preceding the target word is deaccented and, hence, low in prosodic prominence.

2.2. Stimuli

In total, we constructed 40 experimental sentences, i.e., 20 unique sentences per experimental condition: *L+H** on the word preceding the target word, and *deaccentuation* on the word preceding the target word. The experimental sentences consisted of seven words and the target word (the word for which memory recognition was tested) always occurred in the fourth position after the sentential subject (see example (1)). Target words in the experimental sentences were again presented during the recognition phase, to test whether participants correctly recognized that they had heard these words in the afore presented sentences.

The sentential subject preceding the target word was either manipulated with an *L+H** accent or it was deaccented. It was always bi-syllabic and a proper name. Each proper name appeared only once in the experimental sentences (i.e., 40 different proper names were used, 20 per experimental condition). The proper names were segmentally different and were neither (phonological) minimal pairs nor rhymes. Figure 1 depicts the mean time-normalized f0 contours of the word preceding the target word, the sentential subject, in the conditions *deaccentuation* and *L+H**.

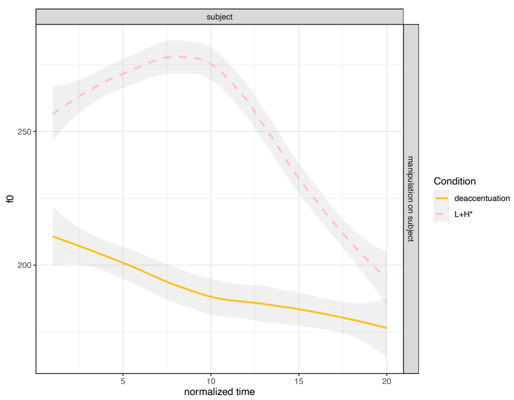

Figure 1: *Mean time-normalized f0 contours of the word preceding the target word in the conditions* deaccentuation *(gold) and* L+H* (*rose*).

The syntactic structure of the experimental sentences was held constant: Adverb-Verb-Subject-target word-Object (see example (1), bold indicating the target word, underlining indicating the word being prosodically manipulated). The word preceding the target word, i.e., the sentential subject that was prosodically manipulated, was in the prenuclear position and the nucleus was sentence-final.

(1) *Gerne trinkt <u>Lola</u>* ***morgens*** *ein Glas Milch.*
Adv V S target word O
'happily drinks Lola in the morning a glass of milk'

All target words in the experimental sentences were bi-syllabic, trochaic adverbials (e.g., *morgens* 'in the morning' in (1)). Each target word only appeared in one experimental sentence (i.e., 40 different target words, 20 per experimental condition). To avoid an influence of the word frequency of the target words on word recognition, we controlled for the word frequency of the target words using WebCelex [12]. The mean word frequency of the target words in the condition *L+H** on the word preceding the target word was 140.2 (*SD* = 288.05), in the condition *deaccentuation*, the mean word frequency of the target words was 141.15 (*SD* = 382.2). The mean difference in word frequency of the target words did not differ significantly across experimental conditions ($t(38)$ = -.009, p = .993). We also controlled for the number of phonemes of the target words. The mean number of phonemes in the condition *L+H** was 6.0 (*SD* = 1.62), whereas it was 6.15 (*SD* = 1.31) in the condition *deaccentuation*. The mean number of phonemes of the target words did not differ significantly across experimental conditions ($t(38)$ = -.322, p = .749). Moreover, the type of adverbial (local, temporal, modal, causal) was matched across the two experimental conditions.

Additionally, prior to running the experiment, we had 23 students rate the written sentences regarding to how natural they perceived the sentences on a 5-point Likert scale from 1, highly natural, to 5, not natural at all. The mean rating in the condition *L+H** was 2.24 (*SD* = .53). In the condition *deaccentuation*, the mean rating was 2.29 (*SD* = .68). The mean ratings of the sentences did not differ significantly across experimental condition ($t(38)$ = -.265, p = .792).

Table 1 presents an overview of the aforementioned characteristics of the target words and experimental sentences across the two experimental conditions.

Table 1: *Characteristics of the experimental conditions.*

	Word frequency	**No. of phonemes**	**Rating**
*L+H**	140.2 (*SD* 288.05)	6.0 (*SD* 1.62)	2.24 (*SD* .53)
deaccentuation	141.15 (*SD* 382.2)	6.15 (*SD* 1.31)	2.29 (*SD* .68)

We furthermore constructed 60 filler sentences. Twenty of the filler sentences followed the syntactic and prosodic structure of the experimental sentences. In those sentences, the target word, an adverbial, also occurred in the fourth position and was tested during the recognition phase. The other 40 filler sentences differed in their syntactic structure from the experimental sentences, they had either an Subject-Verb-Object or an X-Verb-Subject-Object structure. These filler sentences also consisted of seven words. They were recorded naturally as if uttered in spontaneous speech, thus there was no particular prosodic manipulation in the filler sentences. The words presented in the recognition phase for these filler sentences were also bi-syllabic adverbials. These adverbials, however, did not occur in the experimental or filler sentences. Thus, during the recognition memory task participants were expected to respond that the tested words had not been part of the afore presented sentences.

All sentences were recorded during one session in a soundproof cabin by a female trained phonetician, who was a speaker of standard German. The stimuli were recorded in Audacity [13], using a C520 headset and the Scarlett 2i2 3rd generation 2-in, 2-out USB audio interface. The items were cut in PRAAT [14], and loudness was adjusted to -23.0 LUFS using Audacity [13].

The experimental and filler sentences were evenly distributed across ten blocks, each block containing four experimental and six filler sentences. Each block contained two sentences per experimental condition (*L+H** or *deaccentuation* on the word preceding the target word, respectively). The order of the sentences within the blocks was pseudo-randomized, so that two sentences of the same experimental condition did not follow each other. Note that each participant saw each sentence only once. We created two lists that were counterbalanced, i.e., block 1 occurred in the first position in list 1 and in the tenth position in list 2. The order of the sentences within a block stayed the same across lists. Prior to statistical analysis we compared the results of the two lists. As the experimental lists did not differ significantly from each other ($t(52) = -1.631$, $p = .109$), we combined the results of both lists for the remaining statistical analysis.

2.3. Participants

We recruited participants at the University of Cologne. Participation was voluntary and the participants were reimbursed for their participation.

In total, we ran the experiment with 56 participants. We excluded two participants as they reported to be simultaneous bilinguals. This left 54 participants for statistical analysis. All of the remaining participants met the criteria of having German as their first language and having normal or corrected-to-normal visual abilities. All of the participants reported to be neuro-typical. None of the participants reported any hearing or speech disorders. The mean age of the participants was 24 years (range 18–39 years). Eleven participants reported to identify as male, one participant identified as diverse and 42 participants identified as female.

2.4. Procedure

The procedure of the experiment was as follows: First, the participants listened to a block of ten sentences in a row, that were presented via headphones. The sentences were played automatically and after each sentence there was a silence of about 4 seconds. After each block of sentences, ten isolated words successively appeared one by one on the computer screen and participants had to decide for each word successively whether it had occurred in the previously presented block of sentences (recognition phase). Participants had to press a green button with their dominant hand when they thought the word on the screen had occurred in one of the previously presented sentences, or they had to push a pink button with their non-dominant hand when they thought the word had not been part of one of the presented sentences. Words presented during the recognition phase were preceded by a fixation cross (750 ms) and were presented in all caps in the center of the screen. Each word stayed on the screen until the button press, i.e., until the participant had decided whether they had heard the word in the preceding block of sentences. In total, participants had to listen to ten blocks of sentences. After each block (listening and recognition memory task), participants had the possibility to take a self-paced break. Figure 2 displays a visual depiction of the experimental setup.

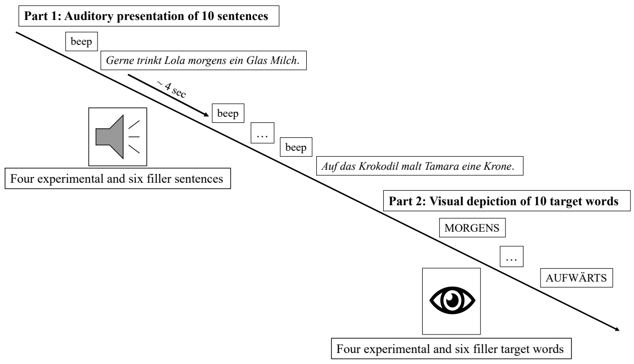

Figure 2: *Visual depiction of the experimental set up.*

The experiment started with an additional practice block of ten practice items. During the practice phase, the participants were able to adjust the loudness to a comfortable level. Participants were randomly assigned to one of the two experimental lists.

Participants were tested individually in a quiet room at the University facilities and were required to wear headphones (Sennheiser HD206) during testing. The experiment was run on a Dell, Latitude 3420 laptop in the open-source program OpenSesame, version 3.3.10. We used the button box by SR Research as a response box.

Prior to the experiment, participants were given oral and written instruction about what they were asked to do in the experiment. Furthermore, they were asked to fill out a questionnaire about their personal and language background, and they gave written informed consent for participation in the experiment.

3. Results

In the condition *deaccentuation* on the word preceding the target word, four data points had to be removed. This was due to an error in the coding of the experiment. We noticed and fixed this error after the fourth participant. No other data points had to be removed. This left in total 1076 data points in the condition *deaccentuation* on the word preceding the target word and 1080 data points for the condition *L+H** on the word preceding the target word. In the condition *L+H** on the word preceding the target word 580 out of these 1080 data points were correct, i.e., participants correctly indicated that they had heard these target words in the preceding block of sentences. This amounts to an accurate recognition rate of 53.7% for the target words in this experimental condition. In the condition *deaccentuation* on the word preceding the target word, 708 out of 1076 data points were correct, this amounts to an accurate recognition rate of 65.8%. Figure 3 shows the mean percentages of correct answers across the two experimental conditions.

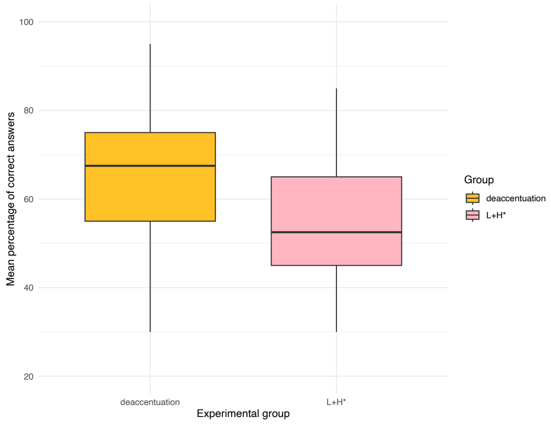

Figure 3: *Plot of the mean percentage of correct answers across experimental conditions.*

The results showed higher recognition rates in the condition *deaccentuation* on the word preceding the target word than in the condition *L+H** on the word preceding the target word. We ran a generalized logistic mixed effects model for the statistical analysis, using the *lme4* package [15] in RStudio [16]. Other than the four data points that had to be excluded due to a programming error, we did not exclude any data points. As the maximally specified model did not converge, we successively reduced the complexity of the model until the model converged. The final model included experimental condition as predictor, random intercepts for items and random intercepts for participants. The model yielded a significant effect of experimental condition, suggesting an effect of prosodic prominence on word recognition memory. Table 2 summarizes the fixed effects of the model.

Table 2: *Fixed effects of the linear mixed effects model.*

	Estimate	**SE**	**z**	**p**
deaccentuation	.752	.173	4.341	
*L+H**	-.579	.217	-2.670	.007

The false alarm rate of the filler sentences that required a "no" as an answer during the testing phase was 19.3%.

4. Discussion

In this study, we examined the effect of prosodic prominence on language processing, specifically on word recognition memory. Critically, unlike previous studies, we did not examine the effect of prosodic prominence on the target word itself but rather examined the effect of prosodic prominence on the word *preceding* the target word. To this end, we conducted a word recognition memory study where the word preceding the target word, for which recognition memory was tested, was prosodically manipulated. We tested two conditions, one in which the word preceding the target word carried a prosodically prominent pitch accent, *L+H** and a condition where the word preceding the target word was deaccented. We found that the target word was recognized less accurately when it followed a word that was marked by a high prominent accent type (*L+H**) compared to when the target word was preceded by a word that was deaccented, i.e., low in prosodic prominence. Thus, our findings show that prosodic prominence can impede the processing of and the memory for a subsequent word, suggesting that effects of prosodic prominence can also have a hindering effect on language processing.

Previous research has shown that prosodic prominence facilitates word recognition memory when manipulated on the target word itself (see e.g., [9], [10]). Higher recognition rates were, for instance, attributed to facilitations in lexical processing. Our findings, however, indicate that the effects of prosodic prominence are not necessarily facilitating. Rather, we would like to suggest that prosodic prominence draws attention to the prominent entity. This leads to a better anchoring of the prosodically prominent word in memory and furthers better recognition rates in recognition memory studies where the target word itself is prosodically prominent as in ([9], [10]). Crucially, however, prosodically prominent entities also bind processing resources by drawing attention and by parsing the prosodic information. This is likely to impede the processing of subsequent sentence material and, consequently, the recognition of sentence material that follows on the prosodically prominent entity, as shown in our study. This proposal can reconcile previous findings with the observations of our present study. In sum, while prosodic prominence can have a facilitatory effect on language processing, we argue that it can also exert a hindering effect on subsequent linguistic material as processing resources are still bound to the processing of the prosodically prominent entity.

The accent *L+H** is not only prosodically prominent but also signals (contrastive) focus. We do not consider these two properties to be mutually exclusive but to interact: while the *L+H** draws attention, it simultaneously evokes the search for contrast alternatives. We argue that both mechanisms result in poorer recognition rates of the following target word, as the attention is bound to the prosodically prominent, contrast evoking entity marked with an *L+H**.

Note that we tested the recognition memory of a target word that *directly* followed the prosodically prominent word. In this setting, we observed a local hindering effect that affects the immediately following word. An interesting issue to pursue in future research is if a target word that does not follow directly on a prosodically prominent word but is separated by more intervening language material from the prosodically prominent word would also be affected from this hindering effect. At this point we have to leave open how local the hindering effect of prosodic prominence is.

In conclusion, our study served to shed new light on the effects of prosodic prominence on sentence processing and word recognition memory. We found that the processing of prosodic prominence is not exclusively facilitative in nature but that it can also be costly when it comes to the processing of information that directly follows the prosodically prominent entity. We suggest that prosodic prominence serves to draw attention to the prominent entity, leading to better recognition memory of the prominent entity, but also binding processing resources that affect the processing of subsequent sentence material.

5. Acknowledgements

We would like to thank Alicia Janz for recording our stimuli and Christine Röhr for her help in choosing a coherent set of stimuli. This research was funded by *Deutsche Forschungsgemeinschaft* (DFG, German Research Foundation) – Project-ID 281511265 – SFB 1252 "Prominence in Language".

6. References

[1] Cutler, A. and M. Pearson. “On the analysis of prosodic turn-taking,” in C. Johns-Lewis (ed.) *Intonation in Discourse*, Routledge, pp. 139–155, 1986.

[2] Schaffer, D. “The role of intonation as a cue to turn taking in conversation,” *Journal of Phonetics*, vol 11(3), pp. 243–245, 1983.

[3] Baumann, S. and M. Grice. “The intonation of accessibility,” *Journal of Pragmatics*, vol 38, pp. 1636–1657, 2006.

[4] Bolinger, D. *Intonation and its part: Melody in spoken English*. Stanford University Press. 1986.

[5] Mücke, D. and M. Grice. “The effect of focus marking on supralaryngeal articulation – Is it mediated by accentuation?,” *Journal of Phonetics*, vol 44(1), pp. 47–61, 2014.

[6] Gussenhoven, C. “Focus, mode and the nucleus,” *Journal of Linguistics*, vol 19(2), pp. 277–417, 1983.

[7] Terken, J. and D. Hermes. “The perception of prosodic prominence,” in M. Horne (ed.) *Prosody: Theory and Experiment*, Kluwer, pp. 89–127, 2000.

[8] Baumann, S. and C. Röhr. “The perceptual prominence of pitch accent types in German,” *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, August 2015.

[9] Fraundorf, S., Watson, D. and A. Benjamin. “Recognition reveals just how CONTRASTIVE contrastive accenting really is,” *Journal of Memory and Language*, vol 63(3), pp. 367–386, 2010.

[10] Kember, H., Jiyoung, C., Yu, J. and A. Cutler. “The processing of linguistic prominence” *Language and Speech*, vol 64(2), pp. 413–436, 2021.

[11] Savino, M., Winter, B., Bosco, A. and M. Grice. “Intonation does aid serial recall after all,” *Psychonomic Bulletin and Review*, vol 27(2), pp. 366–372, 2020.

[12] Baayen. H., Piepenbrock, R. and L. Gulikers. *The CELEX lexical databse (WebCelex)*. University of Philadelphia: Linguistic Data Consortium, 1995.

[13] Audacity Team. *Audacity®: Free Audio Editor and Recorder*. Version 2.4.2, 2021.

[14] Boersma, P. and D. Weenink. *Praat: Doing phonetics by computer*, Version 6.1.16, 2019.

[15] Bates, D., Mächler, M., Bolker, B. and S. Walker. „Fitting linear mixed-effects models using lme4,” *Journal of Statistical Software*, vol 67(1), pp. 1–48, 2015.

[16] R Core Team. *R: A language and environment for statistical computing*, Version 1.4.1106, 2021.