Speech Prosody 2026
26-29 May 2026, Philadelphia, Pennsylvania, USA

What kind of informativity could drive phonetic duration modification: Focus structure or semantic likelihood?

Ivan Yuen, Bernd Möbius, Bistra Andreeva, Mitko Sabev

Saarland University
{ivyuen, bmoebius, andreeva, msabev}@lst.uni-saarland.de

Abstract

Prosody is used to signal 'informativity', which has been separately approached in terms of information structure or information theory. As pointed out in [1], few studies combined both viewpoints in examining prosodic encoding. [1] used meaning-based contextual probability as an information-theoretical measure in an experiment and reported its influence on the fundamental frequency contour in different focus conditions in American English. A recent study of broadcast data in German also observed contributions of information status and trigram surprisal (i.e. structure-based) on syllable duration [2]. Inspired by [1], the current study revisited the role of information structure and information theory on prosodic encoding in German, by using a reading-aloud production experiment, and a meaning-based information-theoretic measure (i.e. likelihood of semantic association between two nouns) as in [1]. We hypothesized that (1) a focused component will exhibit longer duration than its non-focused counterpart, (2) a repeated word will have short duration and (3) a semantically less likely N1-N2 pairing will attenuate their durational differences and therefore attenuate the prominence relationship in each focus structure. Based on data from 10 participants, our preliminary findings provide support for (1), partial support for (2) but no support for (3).

Index Terms: prosodic encoding, information structure, semantic probability, focus, duration

1. Introduction

Information is not distributed equally in a sentence, with some parts being more informative than others. Prosodic encoding is used to signal informativity. The prosody-informativity relation has been mostly investigated from the perspective of information structure, for example, [3] and [4] or information theory, for example, [5], [6] and [7]. According to the information structure framework, an informative message will be emphasized and its phonetic correlates will be associated with higher fundamental frequency (f0), higher intensity, longer duration and exaggerated spectral characteristics. According to the information theoretical framework, a less informative message will be more predictable and its phonetic correlates will be associated with lower f0, less intensity, shorter duration or reduced spectral characteristics. Of relevance is one of the few studies examining both [1]. In [1] a production experiment was carried out to examine the role of (a) information structure status, (b) lexical frequency of nouns and (c) contextual probability of two nouns on prosodic encoding of a sample sentence in American English, such as '*They found fish in the sea*'. They used prompt questions to manipulate 3 information structures of the phrase '*fish in the sea*': Broad focus, Narrow new information focus and Narrow corrective focus. Contextual probability was manipulated by changing the phrase into '*shells in the sea*'. Their analysis of f0 pattern showed that the information structure manipulations affected primarily the post-focal region (i.e. after the object noun). Interestingly, they also reported selective combined effects of lexical frequency and contextual probability. That study suggests both information structure and information theoretic influences on prosodic encoding.

A recent study of broadcast corpus in German also found contributions from surprisal and information status to duration modifications [2], in line with [1] and [8]. However, [2] differed from [1] in a number of ways: the former is corpus-based and uses trigram surprisal (which is structure-related) as an index of information-theoretical measure; whereas the latter is experiment-based and uses pragmatic contextual probability (which is discourse meaning-related). Inspired by [1], the current study revisits the questions in [2] about the roles of information-structural factors and information-theoretical factors on prosodic encoding in German by carrying out a production experiment and manipulating a meaning-based information-theoretic measure, i.e., semantic likelihood of two nouns, as in [1]. Since most of the evidence for information-theoretic measures is based on duration, for example, [5], [6], [7] and [8], the current study will use duration as the measurement variable in light of the findings in [2].*Predictions*: It has been observed that duration is one of the phonetic cues used to distinguish different focus types in German, for example, [4]. In light of that, we predict (1) a narrow-focused component to exhibit longer duration than its broad-focused counterpart in accordance with the information structure perspective.

Some words in prompt questions to elicit different focus conditions are 'given' information that a participant will include in formulating a response. For instance, PROMPT QUESTION -Was hat Rita in der Kirche gefunden? *(What did Rita find in the church?)* and RESPONSE - Sie hat Bibel in der Kirche gefunden (*She has found a Bible in the church*). The phrase 'in der Kirche' is shared in PROMPT QUESTION and RESPONSE. Those 'shared/given' words are analogous to 'predictability'. According to the information-theoretical account, we predict (2) they will be prone to shortening as in [8].

In light of observations in [1] on f0 patterns, we expect similar behaviour in durational patterns. That is, we predict that (3) semantic likelihood manipulation will attenuate the phonetic differences between the two nouns and consequently reduce the prominence relation between the two nouns in each focus condition.

10.21437/SpeechProsody.2026-109

2. Method

A production experiment was carried out in a sound-attenuated booth, using a reading aloud task in question-answer format.

2.1. Stimuli

There were 4 Object nouns to be associated with one of the 8 Place nouns to construct 32 pragmatically plausible, and thus semantically congruent, test Object-Place pairs as shown in Table 1. Each pair was used to construct a target phrase: N1 (Object) + preposition + N2 (Place). The target phrase occurred utterance-medially. All object nouns contained 2 syllables with primary stress on the first syllable. Place nouns contained syllables ranging from 1 to 3 syllables in length. This could not be avoided because we chose to prioritize semantic association/expectation between the nouns while keeping the lexical frequency comparable.

Table 1: *Stimuli containing pragmatically likely/congruent object-place (N1-N2) association*

Object (N1)	Place (N2)
Kühlschrank, Sessel, Fliesen, Dusche *(fridge, chair, tiles, shower)*	Haus *(house)*
Reifen, Lenkrad, Bremse, Blinker *(tyres, steering wheel, brakes, windscreen)*	Auto *(car)*
Tacker, Locher, Drücker, Schreibtisch *(stapler, puncher, printer, desk)*	Büro *(office)*
Schinken, Früchte, Kasse, Kunden *(ham, fruit, cashier, customers)*	Supermarkt *(supermarket)*
Liege, Maske, Pille, Spritze *(lounger, masks, pills, syringe)*	Arztpraxis *(clinic)*
Glocken, Bibel, Predigt, Orgel *(bells, bible, sermon, organ)*	Kirche *(church)*
Filme, Leinwand, Popcorn, Vorhang *(film, screen, popcorn, curtain)*	Kino *(cinema)*
Blume, Vögel, Bänke, Brunnen *(flowers, birds, benches, fountain)*	Park *(park)*

Incongruent Object-Place pairs were then constructed by re-associating the Object nouns with a different and semantically less likely/expected Place noun. This resulted in another 32 less likely Object-Place pairs. The 64 target pairs (inclusive of more and less likely versions) were embedded in a carrier sentence to be read aloud in 3 focus conditions: (1) Broad focus on both Object and Place; (2a) Narrow focus on Object; (2b) Narrow focus on Place. This resulted in a total of 192 target sentences. Each sentence was displayed orthographically in one line as exemplified in the samples below:

Likely Object (N1) -Place (N2) pair

Sie hat eine **Bibel in der Kirche gefunden.**

(*She has found a bible in the church.*)

Less likely Object-Place pair

Sie hat eine **Bibel in der Arztpraxis** gefunden.

(*She has found a bible in the medical practice.*)

Another 106 non-target items (referred to as distractors) were added to the mix. To avoid fatigue and keep the experiment within an hour, the full set (298 sentences) was halved to create two versions. Participants were assigned to each version alternately.

2.2. Participants

Twenty monolingual German-speaking participants took part in the reading aloud experiment, of which 10 were examined in this paper. The average age of this sub-group was 27.6 years in the range of 19 to 51 years, with 6 female and 4 male speakers. Six speakers were assigned to version A, and 4 to version B.

2.3. Procedures

Participants stood in front of a standing microphone which was placed to the side of their mouth corner and individually adjusted to the participant height. A display screen was placed in front for viewing. Test sentences were visually displayed on the screen for reading aloud. In each trial participants first heard a prompt question intended to elicit different focus conditions, followed by an orthographically presented target sentence, which stayed on-screen for 5 seconds. There were 3 types of prompt questions: (1) **Was hat Rita gemacht?** *(What did Rita do?)* for Broad focus elicitation, (2a) **Was hat Rita in der Kirche gefunden?** *(What did Rita find in the church?)* for Narrow focus (Object) elicitation, (2b) **Wo hat Rita die Bibel gefunden?** *(Where did Rita find the bible?)* for Narrow focus (Place) elicitation. Trials were automatically advanced. Three non-test items were used for practice to familiarize the participants with the task and trial structure prior to testing. The experiment was presented using Labvanced[1]. Recordings were made in a sound-attenuated studio using Audacity at a sampling rate of 44.1 kHz with 16-bit quantization.

2.4. Annotation

Since our stimuli contained different segment types, our annotation was based on the following principles. We used auditory, spectrographic and waveform information as a guide. To identify the onset or coda consonant of N1 and N2 in the target phrase, we used the burst release for stops, and the onset of frication for fricatives. For few items containing /l/ and /r/ and nasals, we used lateral release (if present), weak/attenuated spectral amplitude arising from the presence of a secondary resonance cavity for laterals, rhotics and nasals, as well as anti-formants for nasals. Acoustic differences from neighbouring segments were also considered to determine the boundary. If N1 or N2 begins or ends with a vowel, we used F1 and F2 as indicators. In case of contiguous vowels, we used the turning point in formants and change in the shape of waveform as a guide. These principles were applied to delimit the target phrase, as well as N1 (Object) and N2 (Place) in the target phrase. A sample annotation of the target phrase is illustrated in Figure 1.

[1]https://www.labvanced.com/

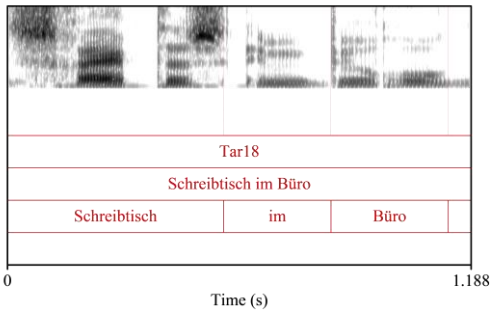

Figure 1: *Sample annotation of the target phrase 'Schreibtisch im Büro' (desk in the office)*

3. Analysis

Durations of the Object and Place nouns in each target phrase were extracted from the annotation in Praat [9]. Separate analysis was carried out in R [10] using raw duration, and durational ratio as a dependent variable. Figure 2 shows the raw mean word duration of Object (N1) and Place (N2) in 3 different focus conditions. Linear mixed effect modelling [11, 12] was used to investigate any effects of FOCUS condition, SEMANTIC LIKELIHOOD condition, and their possible interactions (if modelling permits). Dummy coding was applied to FOCUS, setting Broad focus as the reference level against which Narrow focus (Object) and Narrow focus (Place) were compared, as well as to SEMANTIC LIKELIHOOD with 'more likely' to be the reference level. Speakers and items were included as random effects. The best-fitting model without convergence or singularity issues was identified on the basis of AIC values. The Satterthwaite approximation was used to estimate the degrees of freedom for significance testing. 937 sentences were analyzed, after excluding omission, disfluency and incorrect production (n = 23). Overall, models with interaction did not perform better than those without. Therefore, the following reports were based on the results from the latter.

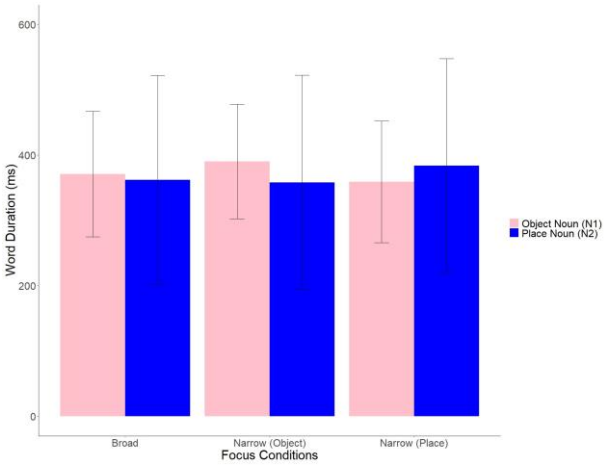

Figure 2: *Mean word duration (ms) of N1 (Object) in pink and N2 (Place) in blue from a target phrase such as 'Schreibtisch (N1) im Büro (N2)' in 3 focus conditions: Broad focus (far left), Narrow focus on N1 (Object) and Narrow focus on N2 (Place) in response to respective prompt questions: What happened? What did she find in the church? Where did she find the Bible?*

3.1. Raw N1 (Object) Duration

The results showed an effect of FOCUS ($F = 52.3$, $df = 2$, $p < .001$*), with Narrow focus (Object) and Narrow focus (Place) both significantly different from Broad focus. As expected, N1 duration is longer in Narrow focus (Object) condition than its counterpart in Broad focus condition in line with the prediction from the information-structure account. No effect of SEMANTIC LIKELIHOOD was observed ($F = 1.25$, $df = 1$, $p = .26$).

Of note is that the estimate for Narrow focus: Place has a negative sign; whereas the estimate for Narrow focus: Object has a positive sign (Table 2). This suggests that N1 duration is shorter in Narrow focus (Place) condition than its counterpart in Broad focus condition. This durational pattern is not expected and will be discussed further in section 4.

Table 2: *Results of the lmer model (~ FOCUS + SEMANTIC LIKELIHOOD + (1|Speaker) + (1|Item) on NP1 duration.*

Fixed effects	Est.	df	t	p-value
Intercept: Broad focus	369.3	18	18.4	*<.001**
Narrow focus: Object	19.9	891	6.2	*<.001**
Narrow focus: Place	-12.7	891	-3.9	*<.001**
Less likelihood	2.9	891	1.1	.26

3.2. Raw N2 (Place) Duration

The results revealed a significant effect of FOCUS ($F = 28.4$, $df = 2$, $p < .001$*) and SEMANTIC LIKELIHOOD ($F = 5.13$, $df = 1$, $p = .024$). As expected, N2 duration is longer in Narrow focus (Place) than its counterpart in Broad focus. Unlike the observations in N1, N2 duration is subject to the effect of SEMANTIC LIKELIHOOD, yet the direction is contrary to expectation. N2 duration is shorter in 'less likely' Object-Place pairs (as reflected in the negative sign of the estimates in Table 3).

Table 3: *Results of the lmer model (~ FOCUS + SEMANTIC LIKELIHOOD + (1|Speaker) + (1|Item) on NP2 duration.*

Fixed effects	Est.	df	t	p-value
Intercept: Broad focus	362.5	14.5	8	*<.001**
Narrow focus: Object	-6.3	917	-1.7	.1
Narrow focus: Place	20.6	917	5.5	*<.001**
Less likelihood	-6.9	917	-2.2	*.02**

3.3. Relative Durational Ratio (N1/N2)

Since Place noun (N2) varied in the number of syllables from 1 to 3, it could affect the ratio measure. Therefore, we included WORD LENGTH OF PLACE NOUN as an additional factor (with dummy coding) in our analysis. The results of the lmer model showed an effect of FOCUS ($F = 66.9$, $df = 2$, $p < .001$*) and

WORD LENGTH OF PLACE NOUN (F = 53.5, df = 2, p < .001*). No effect of SEMANTIC LIKELIHOOD was observed.

Table 4: *Results of the lmer model (~ FOCUS + SEMANTIC LIKELIHOOD + WORD LENGTH OF PLACE NOUN + (1|Speaker) + (1|Item) on NP1/ NP2 ratio.*

Fixed effects	Est.	df	t	p-value
Intercept: Broad focus	1.5	71	20.5	<.001*
Narrow focus: Object	.1	866	5.8	<.001*
Narrow focus: Place	-.1	865	-5.8	<.001*
Less likelihood	.06	64	.9	.38
Word length: 2	-.3	64	-4.2	<.001*
Word length: 3	-.9	64	-10	<.001*

As expected, the N1/N2 ratio is larger in Narrow focus (Object) than its counterpart in Broad focus. Consistently, the N1/N2 ratio is smaller in Narrow focus (Place) than its counterpart in Broad focus. Likewise, the effect of WORD LENGTH OF PLACE NOUN is consistent with expectation. Words containing more syllables have a smaller ratio (as reflected in Table 4 by the negative sign in the estimates).

3.4. Relative Normalized Duration Ratio (raw N1 duration/ average syllable duration in N2)

To de-confound the effect of WORD LENGTH OF PLACE NOUN on the ratio measure, we employed a different relative metric which took word length of N2 (Object) into consideration before calculating the N1/N2 ratio. In other words, we calculated the average syllable duration of N2 before deriving the ratio between N1 and N2. In this analysis, the model did not include WORD LENGTH OF PLACE NOUN as a factor. The model results showed a significant main effect of FOCUS (F = 73.5, df = 2, p <.001*). No effect of SEMANTIC LIKELIHOOD was observed. As predicted, the ratio is larger in Narrow focus (Object) condition than Broad focus condition.

Table 5: *Results of the lmer model (~ FOCUS + SEMANTIC LIKELIHOOD + (1|Speaker) + (1|Item) on normalized duration ratio.*

Fixed effects	Est.	df	t	p-value
Intercept: Broad focus	2.1	72	18.4	<.001*
Narrow focus: Object	.2	866	6.2	<.001*
Narrow focus: Place	-.2	865	-5.9	<.001*
Less likelihood	.05	64	.3	.8

4. Discussion & Conclusions

In using a different method to investigate the roles of information structure and information theory on prosodic encoding in German, our findings revealed a robust effect of information structure as indexed by FOCUS, but a weak and selective effect of an information theoretic measure as indexed by SEMANTIC LIKELIHOOD. Overall, our preliminary findings are similar to findings based on German corpus data in [2] where the information theoretic measure is based on trigram surprisal. It seems that both structure-based and semantic-based 'surprisal' can influence phonetic encoding. However, this is not manifest in the relative metric, suggesting no evidence of prediction (3) that semantic likelihood will affect the phonetic differences between the two nouns and thus reduce the prominence relation between the two nouns in each focus condition.

In our paradigmatic comparisons, the focused component is conveyed by longer duration than its non-focused counterpart. Given the argumentation of the need for a comparison of focused and non-focused components within a phrase at the syntagmatic level [13], our syntagmatic analyses using a duration ratio further support the observations from the paradigmatic comparisons, in line with prediction (1).

Interestingly, we observed two unexpected duration patterns. First, N1 duration in the Narrow focus on Place condition is significantly shorter than its counterpart in the Broad focus condition in the paradigmatic comparison. Since the Object noun is mentioned in the prompt question to elicit the Narrow focus (Place), its 'givenness' status and therefore its predictability could have led to shortening. This observation is in line with the information-theoretical framework, as predicted (2). However, this prediction is not found in N2. Therefore, prediction (2) is only partially borne out.

Second, the effect of SEMANTIC LIKELIHOOD is only observed in N2, but not N1. Although these observations partially support the idea that information theoretical measures could influence prosodic encoding, the direction of influence is contrary to our expectation. In semantically less likely N1-N2 (Object-Place) pairs, its effect is to shorten N2 duration. In other words, unpredictable or improbable pairing does not necessarily induce lengthening or accentuating the durational difference between N1 and N2, as reflected in the lack of a SEMANTIC LIKELIHOOD effect in our ratio analysis.

Our current observations could be limited by using a sub-set of the data, and restricting to the durational analysis. Further analysis of the full data set and other dimensions such as f0 will shed light on the connection between information structure and information theory on prosodic encoding.

To sum, our study showed that both information structure such as focus and information-theoretic measures such as semantic likelihood between N1 and N2 could selectively modulate word duration. These findings are in line with those observed in [1] and [8] in American English and [2] in German.

5. Acknowledgement

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project ID 232722074-SFB1102.

6. References

[1] I. Ouyang and E. Kaiser, "Prosody marks different kinds of informativity: interactions between frequency, probability and focus," *University of Pennsylvania Working Papers in Linguistics*, vol. 21, issue 1, article 24. 2015.

[2] I. Yuen, B. Andreeva, O. Ibrahim and B. Möbius, "Prosodic factors do not always suppress discourse or surprisal factors on word-final syllable duration in German polysyllabic words," in R. Lemke, L. Schäfer, and I. Reich editors, *Information Structure and Information Theory*, pp. 215-234. Language Science Press, Berlin. 2024.

[3] J. Cole, “Prosody in context: a review,” Language, Cognition and Neuroscience,” 30(1-2): 1-31. 2015.
[4] S. Baumann, M. Grice and S. Steindamm, “Prosodic marking of focus domains – categorical or gradient?” *Proceedings of Speech Prosody*, Dresden. 301-304. 2006.
[5] M. Aylett and A. Turk, “The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence duration in spontaneous speech,” *Language and Speech,*. 47: 31-56, 2004.
[6] M. Aylett and A. Turk, “Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei,” *Journal of the Acoustical Society of America,* 119 (5): 3048-3058. 2006.
[7] S. Seyfarth, “Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation,” *Cognition,* 133:140-155. 2014.
[8] R. Baker and A. Bradlow, “Variability in word duration as a function of probability, speech style and prosody,” *Language and Speech*, 52(4): 391-413. 2009.
[9] P. Boersma and D. Weenink, Praat: doing phonetics by computer [Computer program].
[10] R Core Team, *R: A Language and Environment for Statistical Computing*. *R Foundation for Statistical Computing*, Vienna, 2023.
[11] D. Bates, M. Maechler, B. Bolker, and S. Walker, “Fitting Linear mixed-effects models using lme4,” *Journal of Statistical Software,* vol. 67, pp. 1-48, 2015.
[12] R. Lenth, “emmeans: Estimated Marginal Means, aka Least-Squares Means,” *R package version 1.7.4-1,* 2022.
[13] J. Katz and E. Selkirk, “Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English,” *Language,* 87:771-816.