Chapter 8

Prosodic factors do not always suppress discourse or surprisal factors on word-final syllable duration in German polysyllabic words

Ⓘ Ivan Yuen, Ⓘ Bistra Andreeva, Ⓘ Omnia Ibrahim &
Ⓘ Bernd Möbius
Saarland University

Predictability is known to influence acoustic duration (e.g., Ibrahim et al. 2022) and prosodic factors such as accenting and boundary-related lengthening have been postulated to account for this effect (e.g., Aylett & Turk 2004). However, it has also been shown that other factors such as information status or speech styles could contribute to acoustic duration (e.g. Baker & Bradlow 2009). This raises the question as to whether acoustic duration is primarily subject to the influence of prosody that reflects linguistic structure including predictability. The current study addressed this question by examining the acoustic duration of word-final syllables in polysyllabic words in DIRNDL, a German radio broadcast corpus (e.g. Eckart et al. 2012). We analysed polysyllabic words followed by an intermediate phrase or an intonational phrase boundary, with or without accenting, and with given or new information status. Our results indicate that the acoustic duration of the word-final syllable was subject to the effect of prosodic boundary for long host words, in line with Aylett & Turk (2004); however, we also observed additional effects of information status, log surprisal and accenting for short host words, in line with Baker & Bradlow (2009). These results suggest that acoustic duration is subject to the influence of prosodic (e.g., boundary and accenting) and linguistic factors (e.g., information status and surprisal), and that the primacy of prosodic factors impacting on acoustic duration is further constrained by some intrinsic durational constraints, for example word length.

Ivan Yuen, Bistra Andreeva, Omnia Ibrahim & Bernd Möbius

1 Introduction

Information-theoretic measures have been used to account for variabilities in word length (e.g., Piantadosi et al. 2011), phrase duration (e.g., Arnon & Cohen-Priva 2013), word duration (e.g., Baker & Bradlow 2009, Seyfarth 2014), syllable duration (e.g., van Son & Pols 2003, Aylett & Turk 2004), vowel spectra (e.g., Aylett & Turk 2006, Brandt et al. 2021), vowel and consonant dispersion (e.g., Malisz et al. 2018), and lenition (e.g., Cohen-Priva 2017). These studies suggest that speakers' choice of phonetic forms is guided by informativity-based considerations, which include frequency (e.g., Gahl 2008, Arnon & Cohen-Priva 2013), or contextual predictability (e.g., Aylett & Turk 2004, 2006, Baker & Bradlow 2009, Seyfarth 2014, Piantadosi et al. 2011).

Adopting the information-theoretic perspective, Aylett & Turk (2004) postulated the smooth signal redundancy hypothesis (SSRH) to explain the acoustic variability of duration in English. According to this hypothesis, prosody directly affects speech acoustics through assignment of prosodic prominence or boundary. For the sake of robust optimal communication, this prosodic influence is inversely related to the influence from language predictability/redundancy factors in order to maintain smooth transmission of information (i.e., to avoid any abrupt surge or dip in information density). While language predictability will induce short acoustic duration, prominence will induce long acoustic duration. For instance, *nine* in *a stitch in time saves nine* will have shorter duration than the same word in *the winning number is nine*, because nine is more predictable and less prominent in the former than the latter. To test their hypothesis, they analysed syllable duration in the HCRC Map Task Corpus (Anderson et al. 1991), and found a significant inverse relationship between the language predictability factors (e.g., log word frequency, syllabic trigram probability and word mention) and syllable duration. They also observed significant influences on syllable duration from a range of prosodic factors (e.g., lexical stress, phrasal stress, different types of prosodic boundary). Further regression analysis revealed that the model with prosodic prominence structure accounted for most of the variance in the syllable duration, with little unique significant contribution from the language predictability factors. On such basis, they argued that prosody absorbs the effects from language predictability to influence speech acoustics. In other words, prosody mediates language predictability.

However, other studies showed that predictability can directly influence duration, rather than be mediated through prosodic prominence structure. For instance, Baker & Bradlow (2009) examined two predictability factors on reduction (i.e., word duration) in two speech styles: plain vs. clear. Plain style was defined as

one in which a speaker hypo-articulates because listeners do not have difficulty in perceiving one's speech; and clear style as one in which the speaker hyper-articulates because listeners might have difficulty in perceiving one's speech. The two predictability factors were first vs. second mention, and word frequency. As expected, word duration in clear speech is longer than that in plain speech for first and second mentions. Similarly, word duration is longer for first mention than second mention in both speech styles. These patterns remained, irrespective of the presence vs. absence of a prosodic break around the measured targets, or the presence vs. absence of accenting on the target stimuli. On the whole, word frequency is negatively correlated with first-mention word duration in plain and clear speech (short duration for high frequency words). However, this frequency-induced reduction effect is exaggerated on second mention duration in plain speech, not clear speech. Baker & Bradlow (2009) found a significant positive correlation between word frequency and second mention in plain speech (suggesting high frequency words undergo more second mention reduction), but not in clear speech. These findings then support the idea that other non-prosody factors such as discourse structure, lexicon-based frequency (predictability) and speech styles contribute to acoustic realization, with the implication that the connection between predictability and acoustics can be direct, in addition to being mediated through prosody.

Consistent with Baker & Bradlow (2009), a recent study reported independent effects of syllable-based surprisal (one type of predictability) and Lombard speech style (i.e., speech produced in a noisy environment) on syllable duration in German lab-speech (Ibrahim et al. 2022). Further evidence for the effect of surprisal was observed on word-final syllable duration in German preceding an intonational phrase (IP) boundary (Andreeva et al. 2020). Based on the analysis of the DIRNDL corpus (Eckart et al. 2012), the authors showed that the duration of a German word-final syllable with high surprisal was longer than that with low surprisal. Critically, they found an interaction between surprisal and the strength of an IP boundary, with the effect of surprisal being more pronounced in the presence of a strong IP boundary. The presence of such interaction indicates that both prosodic boundary and surprisal contributed to the acoustic duration of the word-final syllable in German. Interestingly, the duration of a word-final syllable preceding a strong IP boundary was shorter than that preceding a weak IP boundary.

However, this study did not differentiate between monosyllabic and polysyllabic words preceding an IP boundary. While the location of lexical stress does not vary for monosyllabic words, this cannot be said for polysyllabic words. The number of syllables may then be confounded with pitch accenting (associated

with lexical stress) in influencing the measured syllable duration. Pitch accenting is typically associated with focus or information status (e.g., Cooper et al. 1985, Cruttenden 1993, 2006), although the distribution and use of accenting can be language-specific (e.g., Swerts et al. 2002). For instance, Swerts and colleagues reported Dutch speakers accenting new and contrastive information, but not given information. Cruttenden (1993) showed that speakers of English de-accent repeated or old information; however, this tendency will be attenuated in the presence of a contrast in the discourse. As such, this raises further questions as to whether accenting or information status, or both, contribute to the measured duration and whether or not such effect(s) will interact with surprisal. Besides, it remains to be seen whether the effect of surprisal continues to be observed for other types of prosodic boundary, say intermediate phrase (ip) boundary.

To better understand how discourse-based structure (i.e., information structure), language predictability (i.e., surprisal), prosody (i.e., presence vs. absence of accenting or prosodic boundary types), and/or their interactions might account for the acoustic variability of duration, the current study used broadcast data from an annotated German corpus (DIRNDL) to examine any effects of information status (an aspect of discourse-based information structure) and syllable-based surprisal on word-final syllable durations adjacent to an intonational or an intermediate phrase boundary.

Given the previous finding from Aylett & Turk (2004), we expected prosody (through boundary-related lengthening or accenting) to largely account for the acoustic duration of word-final syllables, mediating any effects of surprisal or discourse (e.g., information status). However, according to observations from Baker & Bradlow (2009), Andreeva et al. (2020) and Ibrahim et al. (2022), we also expected prosodic, surprisal and discourse factors (or their interactions) to contribute to the acoustic duration.

2 Method

We extracted polysyllabic words from the DIRNDL corpus to empirically test whether information status and syllable-based surprisal moderate word-final syllable duration in two prosodic boundary types in German. The DIRNDL corpus (Discourse Information Radio News Database for Linguistic analysis) consists of 5 hours of audio news recordings in German from 9 speakers (5M, 4F) with prosodic annotations for pitch accent types and boundaries according to the GToBI(S) framework (Mayer 1995). The accompanying written scripts were annotated for information status (see Eckart et al. 2012 for details of corpus construction and segmentation).

2.1 Data selection criteria

A total of 3716 polysyllabic words were identified to occur before either an intermediate phrase (as denoted by ip) or an intonational phrase (as denoted by IP) boundary in the DIRNDL corpus. The word-final syllable constitutes the target syllable because it occurs immediately adjacent to a prosodic boundary. As not all of the identified polysyllabic words before a prosodic boundary were annotated for lexical information status, the data set was further trimmed to include only those with a clearly specified information status. Note that information status was grouped into two levels: given vs. new. Items annotated as "accessible" in the DIRNDL corpus were classified as "given" in the current study. This procedure reduced the data set to a total of 2907 items for statistical analysis. Information related to the host word containing the target syllable were extracted from the DIRNDL corpus: speaker identity, speaker gender, identity of the orthographically transcribed host word, phonemic transcription of the target syllable, pitch accent type (if present) for the host word, prosodic boundary (i.e., intermediate or intonational), and lexical information status.

2.2 Language modelling

We estimated the syllable-based surprisal measure in the current study from language models based on the deWaC (deutsches Web as Corpus) corpus (Baroni et al. 2009). The corpus is a collection of web-crawled data containing about 1.7 billion word tokens and 8 million word types from a diverse range of genres such as newspaper articles and chat messages. The corpus was first pre-processed and normalized using German Festival (Möhler et al. 2000). This procedure consisted of removing unnecessary/irrelevant/duplicate document information, for example, web-specific structures such as HTML structures or long lists. After pre-processing, the normalized corpus was divided into a training set (80%) and a test set (20%). Syllable-based trigram language models including word boundary as a unit were trained on the training set using the SRILM toolkit (Stolcke 2002). All language models underwent Witten-Bell smoothing (Witten & Bell 1991). The best-performing trained language model was then used as the default to calculate the conditional probability of a syllable, given the preceding context, i.e.,

(1) $$S(\text{unit}_i) = -\log_2 P(\text{unit}_i|\text{context})$$

where S = surprisal and P = probability (Hale 2016). The context consisted of two units/states: syllable or/and word boundary. The conditional probability constituted the syllable-based surprisal measure for the target syllables.

2.3 Analysis

Prior to the main analysis, we first checked the estimated surprisal values of the target syllables in the polysyllabic words preceding the two prosodic boundaries and observed two patterns: (a) target syllables preceding an intonational phrase boundary (IP) had overall higher surprisal values than those preceding an intermediate phrase boundary (ip) when the host word contained 2 or 3 syllables, (b) target syllables preceding an intermediate phrase boundary had overall higher surprisal values than those preceding an intonational phrase boundary when the host word contained 4 to 8 syllables. In (a) higher surprisal values were associated with an intonation phrase boundary (IP); whereas in (b) higher surprisal values were associated with an intermediate phrase boundary (ip) (see Figure 1). Because of this, we divided the full data set of 2907 polysyllabic words into two separate data sets to de-confound the effect of surprisal from that of prosodic boundary: 2317 words with no more than 3 syllables and 590 words with no more than 8 syllables (but at least 4 syllables). The former was referred to as *short words*, and the latter as *long words* hereafter. A custom Python script was then used to extract durations of the word-final target syllables from the DIRNDL corpus.
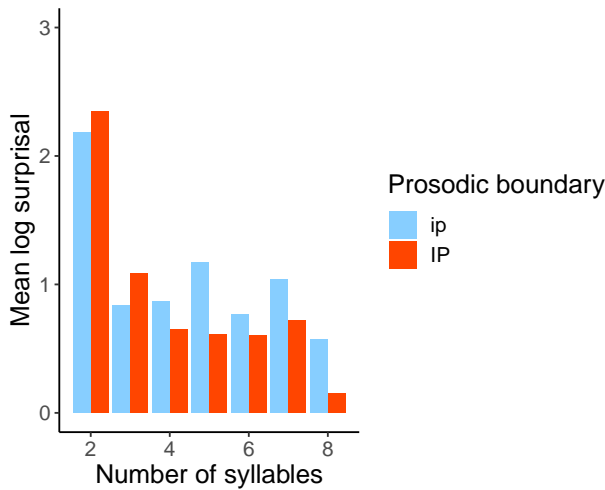

Figure 1: Mean syllable-based log surprisal values of the final syllables according to the number of syllables in the polysyllabic words and prosodic boundary

The duration of the word-final syllable was the dependent variable. Predictors included prosodic boundary type (intermediate vs. intonational), information status (given vs. new), word length (short vs. long), presence vs. absence of a pitch

accent, log surprisal of the word-final syllable, prosodic boundary * information status interaction, presence vs. absence of a pitch accent * information status interaction, log surprisal * prosodic boundary interaction, and log surprisal * information status interaction. Random factors included speaker identity and syllable identity.

3 Results

Linear mixed effects models were then fit to the dependent variable, namely word-final syllable duration, using the lme4 package (Bates et al. 2015) in R (R Core 2022). Multiple random structures were first constructed and compared, using AIC (Alkaike Information Criterion) to determine the optimal random structure as the baseline model. The baseline random structure included by-speaker and by-item intercepts, with prosodic boundary by-speaker slope. Predictors were then included in the baseline model to construct simple and interactive models, which were compared using AIC in order to determine the optimal predictive model. In case of singularity, the complexity of a model structure was reduced to minimize overfitting. Statistical significance of the predictors in the optimal model was then evaluated using the anova() function with Satterthwaite to approximate degrees of freedom. The same procedures were followed in all analyses below. The factors included prosodic boundary, information status, log surprisal of the word-final syllable, presence vs. absence of a pitch accent for the host word, word length, and their interactions. The omnibus analysis revealed significant effects of prosodic boundary, presence vs. absence of a pitch accent, log surprisal, with a significant 2-way prosodic boundary * word length interaction and a significant 3-way prosodic boundary * information status * word length interaction (Table 1). To better understand the 3-way interaction, we analysed short and long words separately.

3.1 Final syllable duration in short words

Figure 2 illustrates the word-final syllable duration in accented or unaccented host words preceding an intermediate or intonational phrase boundary, labelled as having the information status of either "given" or "new". The word-final syllable duration was increasingly longer when the prosodic boundary immediately following the host word became stronger (i.e., intermediate vs. intonational phrase) and this pattern was magnified when the host word was accented, as well as when the host word was labelled as "new" information. Table 2 summarizes the mean and standard deviation of word-final syllable durations in the 2 prosodic

Table 1: Statistical results of linear mixed effects modelling on word-final syllable durations in all polysyllabic words. *The model: ~ prosodic boundary * information status * word length + presence vs. absence of accent * information status + prosodic boundary * log surprisal + information status * log surprisal + (prosodic boundary | speaker) + (prosodic boundary | syllable item)*

Factors	F	df	*p*	
Prosodic boundary (PB)	17	1	<0.001	***
Information status (IS)	2.6	1	0.11	
Log surprisal (S)	46.9	1	<0.0001***	
Word length (WL)	1.61	1	0.21	
Presence vs. absence of pitch accent (PA)	77.9	1	<0.0001***	
PB * IS	1.9	1	0.17	
PB * WL	4.9	1	0.03	*
IS * WL	0.5	1	0.5	
IS * PA	0.5	1	0.5	
PB * S	3	1	0.08	
IS * S	1.1	1	0.3	
PB * IS * WL	7.5	1	0.006	**

boundary (i.e. intermediate vs. intonational) × 2 information status (i.e. given vs. new) × 2 accenting (i.e. with vs. without) conditions.

Linear mixed effects models were then fit to the dependent variable: word-final syllable duration in short words. The factors in the optimal predictive model included prosodic boundary, information status, log surprisal of the word-final syllable, presence vs. absence of a pitch accent for the host word, and prosodic boundary * information status interaction.

The structure of the final model was ~ prosodic boundary * information status + log suprisal + presence vs. absence of pitch accent + (prosodic boundary | speaker) + (prosodic boundary | syllable item), with significant main effects of prosodic boundary, information status, log surprisal, presence vs. absence of pitch accent and the significant prosodic boundary * information status interaction (Table 3).

As expected, the word-final syllable duration was longer preceding an intonational phrase than an intermediate phrase boundary (prosodic boundary effect). It was longer when the host word contained "new" rather than "given" information (information status effect). It was longer when the host word was accented as opposed to unaccented (presence vs. absence of a pitch accent effect). It was
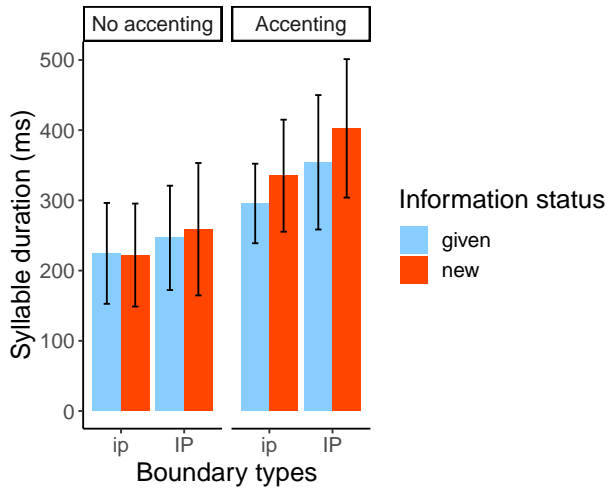
Figure 2: Mean word-final syllable durations in short words, with the following intermediate (ip) or intonational phrase (IP) boundary, with given or new information status, and with or without accent, with +/- 1 SD

Table 2: Mean (SD) word-final syllable durations in short words, with the following intermediate (ip) or intonational (IP) phrase boundary, with given or new information status, and with or without accenting

Accenting	Information status	Prosodic boundary	No. of items	Mean in ms (SD)
No	Given	ip	230	225 (72)
		IP	166	247 (74)
	New	ip	724	222 (73)
		IP	736	259 (94)
Yes	Given	ip	64	296 (57)
		IP	40	354 (96)
	New	ip	173	335 (80)
		IP	184	403 (99)

Table 3: Statistical results of linear mixed effects modelling on word-final syllable duration in all short words. *The model: ~ prosodic boundary * information status + log suprisal + presence vs. absence of pitch accent + (prosodic boundary | speaker) + (prosodic boundary | syllable item)*

Factors	F	df	*p*	
Prosodic boundary (PB)	18.3	1	0.001	**
Information status (IS)	4.3	1	0.04	*
Log surprisal (S)	64.5	1	<0.0001***	
Presence vs. absence of pitch accent (PA)	93.3	1	<0.0001***	
PB * IS	17.7	1	<0.0001***	

also longer when the word-final syllable had high log surprisal (log surprisal effect). However, the significant prosodic boundary * information status interaction suggests that the effect of information status on word-final syllable duration was magnified when the immediately adjacent boundary constitutes an intonational phrase. These results were more in line with the predictions from Baker & Bradlow (2009) than those from Aylett & Turk (2004). Counter to our expectations, log surprisal did not interact with information status or prosodic boundary.

To further investigate this, we sub-divided our data into words with vs. without accenting and separately analysed them. The optimal model for the data with accenting included predictors: prosodic boundary, information status, log surprisal, prosodic boundary * information status interaction, information status * log surprisal interaction, and prosodic boundary * log surprisal interaction. The model structure was ~ prosodic boundary * information status + prosodic boundary * log surprisal + information status * log surprisal + (prosodic boundary | speaker) + (prosodic boundary | syllable item), with log surprisal and information status * log surprisal interaction reaching statistical significance (Table 4). Unlike the analysis of all short words, we observed the effect of log surprisal and the log surprisal * information status interaction.

The optimal model for the data without accenting included the following predictors: prosodic boundary, information status, log surprisal, prosodic boundary * log surprisal interaction, prosodic boundary * information status interaction and information status * log surprisal interaction. The structure of the final model for the data without accenting was ~ prosodic boundary * information status +

Table 4: Statistical results of linear mixed effects modelling on word-final syllable duration in short words with accenting. *The model: ~ prosodic boundary * information status + prosodic boundary * log surprisal + information status * log surprisal + (prosodic boundary | speaker) + (prosodic boundary | syllable item)*

Factors	F	df	*p*	
Prosodic boundary (PB)	0.9	1	0.34	
Information status (IS)	0.7	1	0.4	
Log surprisal (S)	20.4	1	<0.0001***	
PB * IS	2.5	1	0.12	
S * PB	1.4	1	0.24	
S * IS	3.9	1	0.05	*

prosodic boundary * log surprisal + information status * log surprisal + (prosodic boundary | speaker) + (prosodic boundary | syllable item), with significant effects of prosodic boundary, information status, log surprisal and the prosodic boundary * information status interaction (Table 5). These results were consistent with the results in the analysis of all short words, suggesting that the overall pattern might be driven primarily by the data without accenting (which contained more items overall).

Table 5: Statistical results of linear mixed effects modelling on word-final syllable duration in short words without accenting. *The model: ~ prosodic boundary * information status + prosodic boundary * log surprisal + information status * log surprisal + (prosodic boundary | speaker) + (prosodic boundary | syllable item)*

Factors	F	df	*p*	
Prosodic boundary (PB)	25.9	1	<0.001	***
Information status (IS)	5.9	1	0.02	*
Log surprisal (S)	40.2	1	<0.0001***	
PB * IS	13.3	1	<0.001	***

The interaction of log surprisal and information status on word-final syllable duration is illustrated in Figure 3 for accented words and the lack of interaction in Figure 4 for unaccented words. In Figures 3 and 4, the word-final syllable duration was lengthened when the log surprisal value was high, as reflected in the positive correlation. However, in Figure 3, the slope between the word-final

syllable durations and the log surprisal values was conspicuously steeper when host words contained new rather than given information. That is, the effect of surprisal on a word-final syllable duration was attenuated when the host word contained "given" information and accenting. Unlike Figure 3, the differences in the slope between given vs. new information in Figure 4 were less obvious.
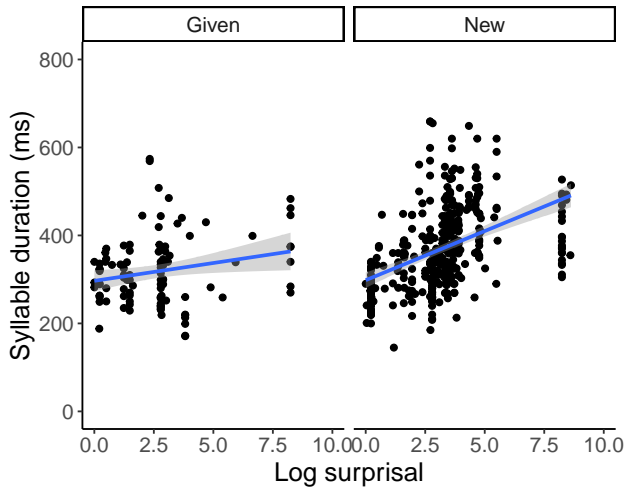

Figure 3: Scatterplots relating word-final syllable durations on the y-axis to syllable-based log suprisal values on the x-axis in accented short host words labelled with given or new information status

3.2 Final syllable duration in long words

A total of 590 long words were included in this analysis. Figure 5 illustrates the patterns of word-final syllable duration preceding two different prosodic boundaries, with either given or new information status, and with or without accenting. Generally, the word-final syllable duration with new information status was longer than that with given information status. This held for unaccented words preceding an intermediate phrase or an intonational phrase boundary; however, the pattern was not as consistent for accented words. Table 6 summarizes the mean duration with SD.

The optimal model included prosodic boundary, information status, log surprisal, presence vs. absence of pitch accent and prosodic boundary * information status interaction as predictors. The model structure of the optimal model was ~ prosodic boundary * information status + log surprisal + presence vs. absence
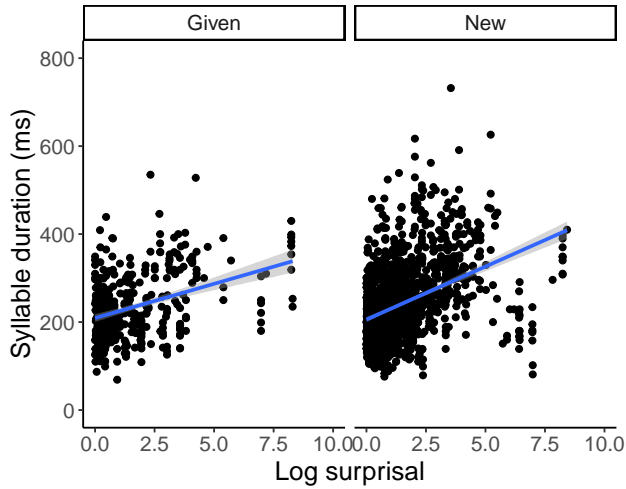
Figure 4: Scatterplots relating word-final syllable durations on the y-axis to syllable-based log suprisal values on the x-axis in unaccented short host words labelled with given or new information status
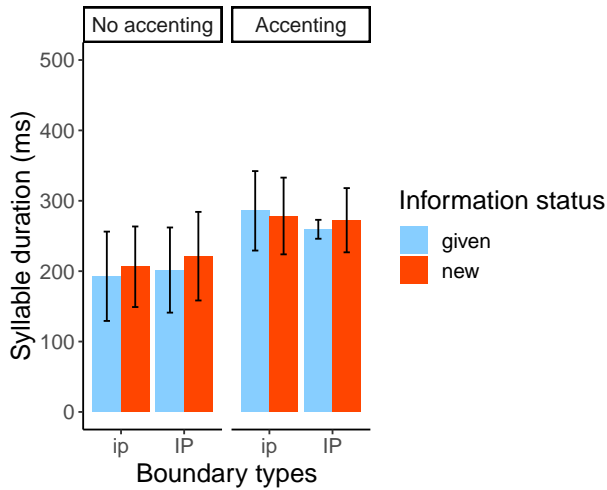

Figure 5: Mean word-final syllable durations in long words, followed by an intermediate (ip) or intonational phrase (IP) boundary, with given or new information status, and with or without accenting, with +/− 1 SD

Table 6: Mean (SD) word-final syllable duration in long words, followed by an intermediate (ip) or an intonational (IP) phrase boundary, with either given or new information status, and with or without accenting

Accenting	Information status	Prosodic boundary	No. of items	Mean in ms (SD)
No	Given	ip	64	193 (63)
		IP	60	202 (61)
	New	ip	232	206 (57)
		IP	213	221 (63)
Yes	Given	ip	4	286 (56)
		IP	2	260 (13)
	New	ip	10	278 (55)
		IP	5	272 (46)

Table 7: Statistical results of linear mixed effects modelling on the final syllable duration in long words. *The model: ~ prosodic boundary * information status + log surprisal + presence vs. absence of pitch accent + (prosodic boundary | speaker) + (1 | syllable item)*

Factors	F	df	p
Prosodic boundary (PB)	15.3	1	0.002**
Information status (IS)	2.3	1	0.13
Log surprisal (S)	0.05	1	0.82
Presence vs. absence of pitch accent (PA)	2.5	1	0.11
PB * IS	2.2	1	0.14

of pitch accent + (prosodic boundary | speaker) + (1 | syllable item), with the significant effect of prosodic boundary (Table 7).

Counter to the prediction from Baker & Bradlow (2009), neither the effect of log surprisal nor its interaction with other predictors were observed on the word-final syllable duration (as exemplified by the lack of log surprisal * information status interaction in Figure 6). These results are more in line with the predictions from Aylett & Turk (2004).
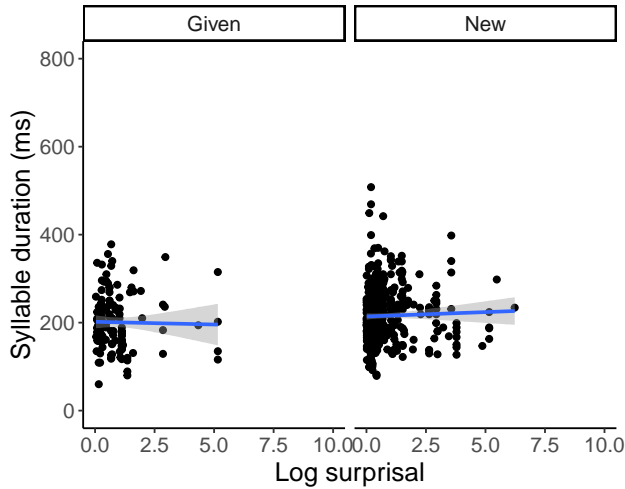
Figure 6: Scatterplots relating word-final syllable durations on the y-axis to syllable-based log surprisal values on the x-axis in all long words with given or new information status

4 Discussion

The goal of the current investigation was to empirically test whether discourse factors such as information status, prosodic factors such as prosodic boundary type, accenting, and surprisal (or their interactions) would contribute to the acoustic realization of the word-final syllable duration in polysyllabic words. We expected the word-final syllable duration with new information status to be longer than that with given information status (e.g., Fowler & Housum 1987, Lam & Watson 2010). We also expected the word-final syllable duration to be longer when followed by an intonational phrase boundary rather than an intermediate phrase boundary (e.g., Wightman et al. 1992). We also expected word-final syllable duration in an accented word to be longer than that in an unaccented word (e.g., Turk & White 1999). We further hypothesized that the word-final syllable duration with high log surprisal would be longer than that with low log surprisal (e.g., Ibrahim et al. 2022). We postulated that prosodic factors would largely account for the acoustic duration without unique contribution from other factors, in line with Aylett & Turk (2004); however, other factors in addition to the prosodic factors might contribute to the measured duration, in line with Baker & Bradlow (2009).

Our overall results on short polysyllabic words are consistent with the effects of boundary-related lengthening, accentual lengthening, surprisal and informa-

tion status on the acoustic realization of syllable duration in previous studies. These results are more in line with Baker & Bradlow (2009) rather than Aylett & Turk (2004), because of additional contributions (including interactions) from prosodic and language predictability factors. However, our overall results on *long* polysyllabic words are more in line with Aylett & Turk (2004) rather than Baker & Bradlow (2009), because we observed only the prosodic boundary effect. Perhaps, polysyllabic shortening constrains the extent to which these various factors can modify syllable duration (i.e., a word length constraint). Despite such a constraint, the durational adjustment is primarily attributable to prosodic boundary type. This suggests a strong tendency for duration to maintain information about a major prosodic boundary. But this interpretation has to be taken with caution because of the low statistical power arising from a relatively smaller data set containing long words (a total of 590 items) than that containing short words (a total of 2317 items).

Depending on whether or not host words were accented, different results were revealed in the data set containing short words. When host words were accented, only the effect of log surprisal and the log surprisal * information status interaction reached statistical significance. However, when host words were unaccented, prosodic boundary, information status, log surprisal and prosodic boundary * information status interaction significantly predicted the word-final syllable durations. As the syllable duration in an unaccented word is shorter than that in an accented word (as reflected in the different y-intercepts in Figure 4 vs. Figure 3), the former might have more room than the latter to accommodate durational increases from multiple sources, resulting in more reported effects and interactions. This interpretation may account for the different results between the unaccented vs. accented words on the assumption that there is an upper duration limit, which seems to be the case, because the majority of the data for word-final syllable duration fell below 600ms (Figure 3 and Figure 4). An alternative interpretation may be related to the statistical power of the relatively small sample size for accented words (461 items) to detect multiple effects as compared to that for unaccented words (1856 items).

Recall that one of our questions is whether information status might interact with the log surprisal effect, presence vs. absence of accenting effect and/or the prosodic boundary effect. Since information status did not have any effect on long words, our discussion focuses on short words. In short *accented* words, information status interacts with log surprisal. The interaction occurs because the effect of log surprisal on word-final syllable duration was stronger for words with new information status than those with given information status. However, no such interaction was observed in *unaccented* words. In that case, information

status interacts with prosodic boundary instead. This interaction is due to the effect of prosodic boundary on pre-boundary syllable duration which is stronger for words with new information than those with given information. In other words, information status additionally exaggerates the effect of log surprisal in accented words, and the effect of prosodic boundary in unaccented words. These results suggest that information status (a discourse-based factor) cannot be subsumed under surprisal (a language predictability factor) or prosodic boundary (a prosodic factor).

These observations based on the corpus data raise further questions as to whether a speaker will differentially weigh these factors (language predictability, information structure and prosody) according to speech styles, e.g., scripted vs. spontaneous or formal vs. informal speech. Speech styles could affect how utterances will be structured, because a speaker might adopt different production planning strategies to cope with time pressure for example. It is possible that less scripted styles might induce more pause breaks. Since these structural considerations can also affect speech acoustics (e.g., Watson & Gibson 2004), it remains to explore in future questions as to how other acoustic cues such as pause might relate to pre-boundary syllable duration.

5 Conclusion

Our overall results showed that information status and surprisal do not encode the same type of linguistic information and that neither information status nor surprisal are redundant. Both can influence the acoustic realization of word-final syllable durations. Critically, our results showed that prosodic factors such as prosodic boundary type could largely account for the acoustic duration as predicted by Aylett & Turk (2004) for long words on the one hand, but that factors other than prosody also contribute to the measured acoustic duration as predicted by Baker & Bradlow (2009) for short words. In other words, information status as a discourse factor can interact with language predictability and prosodic factors to influence the measured acoustic syllable duration, but these interactions are subject to some duration constraint(s) arising from word length.

Acknowledgements

References

Anderson, Anne, Miles Bader, Ellen Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Catherine Sotillo, Henry Thompson & Regina Weinert. 1991. The HCRC map task corpus. *Language and Speech* 34(4). 351–366. DOI: 10.1177/002383099103400404.

Andreeva, Bistra, Bernd Möbius & James Whang. 2020. Effects of surprisal and boundary strength on phrase-final lengthening. In *Proceedings in 10th International Conference on Speech Prosody*, 146–150. DOI: 10.21437/SpeechProsody.2020.

Arnon, Inbal & Uriel Cohen-Priva. 2013. More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech* 56(3). 349–371. DOI: 10.1177/0023830913484891.

Aylett, Matthew & Alice Turk. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47(1). 31–56. DOI: 10.1177/00238309040470010201.

Aylett, Matthew & Alice Turk. 2006. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America* 119(5). 3048–3058. DOI: 10.1121/1.2188331.

Baker, Rachel & Ann Bradlow. 2009. Variability in word duration as a function of probability, speech style and prosody. *Language and Speech* 52(4). 391–413. DOI: 10.1177/0023830909336575.

Baroni, Marco, Silvia Bernardini, Adriano Ferraresi & Eros Zanchetta. 2009. The WaCky Wide Web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43(3). 209–226. DOI: 10.1007/s10579-009-9081-4.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software* 67. 1–48. DOI: 10.18637/jss.v067.i01.

Brandt, Erika, Bernd Möbius & Bistra Andreeva. 2021. Dynamic formant trajectories in German read speech: Impact of predictability and prominence. *Frontiers in Communication* 6. DOI: 10.3389/fcomm.2021.643528.

Cohen-Priva, Uriel. 2017. Not so fast: Fast speech correlates with lower lexical and structural information. *Cognition* 160. 27–34. DOI: 10.1016/j.cognition.2016.12.002.

Cooper, William, Stephen Eady & Pamela Mueller. 1985. Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America* 77. 2142–2155. DOI: 10.1121/1.392372.

Cruttenden, Alan. 1993. The de-accenting and re-accenting of repeated lexical items. In *Working papers*, vol. 41, 16–19. Lund: Department of Linguistics & Phonetics. https://journals.lub.lu.se/LWPL/article/download/17147/15524/.

Cruttenden, Alan. 2006. The de-accenting of given information: A cognitive universal? In Giuliano Bernini & Marcia Schwartz (eds.), *Pragmatic Organization of Discourse in the Languages of Europe*, 311–356. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110892222.311.

Eckart, Kerstin, Arndt Riester & Katrin Schweitzer. 2012. A discourse information radio news database for linguistic analysis. In C. Chiarcos, S. Nordhoff & S. Hellmann (eds.), *Linked data in linguistics: Representing and connecting language data and language metadata*, 65–75. Heidelberg: Springer-Verlag Berlin. DOI: 10.1007/978-3-642-28249-2_7.

Fowler, Carol & Jonathan Housum. 1987. Talkers' signalling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language* 26. 489–504. DOI: 10.1016/0749-596X(87)90136-7.

Gahl, Susanne. 2008. Time and thyme are not homophones: The effect of lemma frequency on word duration in spontaneous speech. *Language* 84(3). 474–496. DOI: 10.1353/lan.0.0035.

Hale, John. 2016. Information-theoretical complexity metrics. *Language and Linguistic Compass* 10(9). 397–412. DOI: 10.1111/lnc3.12196.

Ibrahim, Omnia, Ivan Yuen, Marjolein van Os, Bistra Andreeva & Bernd Möbius. 2022. The combined effects of contextual predictability and noise on the acoustic realization of German syllables. *Journal of the Acoustical Society of America* 152(2). 911–920. DOI: 10.1121/10.0013413.

Lam, Tuan & Duane Watson. 2010. Repetition is easy: Why repeated referents have reduced performance. *Memory and Cognition* 38. 1137–1146. DOI: 10.3758/MC.38.8.1137.

Malisz, Zofia, Erika Brandt, Bernd Möbius, Mi Oh Yoon & Bistra Andreeva. 2018. Dimensions of segmental variability: Interactions of prosody and surprisal in six languages. *Frontiers in Communication* 3(25). DOI: 10.3389/fcomm.2018.00025.

Mayer, Jörg. 1995. Transcription of German intonation: The Stuggart system. In *University of Stuggart Technical Report*. Institute of Natural Language Processing. www.ims.uni-stuttgart.de/documents/arbeitsgruppen/ehemalig/ep-dogil/STGTsystem.pdf.

Möhler, Gregor, Antje Schweitzer, Mark Breitenbücher & Martin Barbisch. 2000. *IMS German Festival (version 1.2-os)*. University of Stuggart: Institut für Maschinelle Sprachverarbeitung (IMS).

Piantadosi, Steven, Harry Tily & Edward Gibson. 2011. Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences* 108(9). 3526–3529. DOI: 10.1073/pnas.1012551108.

R Core, Team. 2022. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.

Seyfarth, Scott. 2014. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition* 133. 140–155. DOI: 10.1016/j.cognition.2014.06.013.

Stolcke, Andreas. 2002. SRILM – an extensible language modeling toolkit. In *Proceedings of interspeech*, vol. 2, 901–904. Denver. DOI: 10.21437/ICSLP.2002-303.

Swerts, Marc, Emiel Krahmer & Cinzia Avesani. 2002. Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics* 30(4). 629–654. DOI: 10.1006/jpho.2002.0178.

Turk, Alice & Laurence White. 1999. Structural influences on accentual lengthening. *Journal of Phonetics* 27(2). 171–206. DOI: 10.1006/jpho.1999.0093.

van Son, Rob & Louis Pols. 2003. Information structure and efficiency in speech production. In *Proceedings of the 8th Conference on Speech Communication and Technology, Eurospeech*. Geneva. DOI: 10.21437/Eurospeech.2003-63.

Watson, Duane & Edward Gibson. 2004. The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes* 19. 713–755. DOI: 10.1080/01690960444000070.

Wightman, Colin, Stefanie Shattuck-Hufnagel, Mari Ostendorf & Patti Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91(3). 1707–1717. DOI: 10.1121/1.402450.

Witten, Ian & Tim Bell. 1991. The zero-frequency problem: Estimating the probabilities of novel events in adaptive text compression. *IEEE Transactions on Information Theory* 37(4). 1085–1094. DOI: 10.1109/18.87000.