Non-uniform cue-trading:
Differential effects of surprisal on pause usage and pause duration in German

Ivan Yuen, Omnia Ibrahim, Bistra Andreeva, Bernd Möbius

Department of Language Science and Technology, Universität des Saarlandes
ivyuen@lst.uni-saarland.de; omnia@lst.uni-saarland.de; andreeva@lst.uni-saarland.de; moebius@lst.uni-saarland.de

ABSTRACT

Pause occurrence is conditional on contextual (un)predictability (in terms of surprisal) [10, 11], and so is the acoustic implementation of duration at multiple linguistic levels. Although these cues (i.e., pause usage/pause duration and syllable duration) are subject to the influence of the same factor, it is not clear how they are related to one another. A recent study in [1] using pause duration to define prosodic boundary strength reported a more pronounced surprisal effect on syllable duration, hinting at a trading relationship. The current study aimed to directly test for trading relationships among pause usage, pause duration and syllable duration in different surprisal contexts, analysing German radio news in the DIRNDL corpus. No trading relationship was observed between pause usage and surprisal, or between pause usage and syllable duration. However, a trading relationship was found between the durations of a pause and a syllable for accented items.

Keywords: surprisal, pause, prosodic boundary, information status, cue-trading

1. INTRODUCTION

Pause serves multiple linguistic or extra-linguistic functions. For instance, it can be used to delimit hierarchical linguistic units [e.g., 24, 27], and is often interpreted as an index to reflect difficulties in word selection [e.g, 10, 11], speech production [e.g., 17] and advance planning [e.g., 14, 15], although it can also express extralinguistic information such as the degree of commitment affecting the recognition of emotions [e.g., 25].

Early research analysing spontaneous speech data has identified predictability as a factor influencing the occurrence of a pause. For instance, Goldman-Eisler [10, 11] found that a pause tended to occur at the juncture (i.e., in the context of) of a lexically-frequent preceding word and a lexically-infrequent following word, compared to their counterparts without any pause. This suggests the usage of pause might be conditioned by contextual predictability. However, using lexical word frequency to define predictability could have confounded semantic and structural factors.

Evidence for the influence of contextual predictability has been accumulated in recent years on the implementation of acoustic cues (e.g., temporal/durational and spectral) at various linguistic levels [phrase: e.g., 2; word: e.g., 21; syllable: e.g., 3, 4, 13; segmental: e.g., 8, 18]. Estimated from trained language models, measures of contextual predictability can go beyond lexical frequency to capture the conditional probability of a target linguistic unit (operationalized in the current study as surprisal).

Although the effect of contextual predictability has been independently investigated on the incidence of a pause and the acoustic realization of duration, no study has attempted to bring them together and examine whether and how these two cues could be related to one another, when both are known to be subject to the influence of contextual predictability.

According to the Smooth Signal Redundancy hypothesis, predictability affects the acoustic realization of duration, with short duration for predictable information, in order to avoid an abrupt surge in information during the transmission of signals [e.g., 3]. Although Aylett and Turk [3] argued that prosody could mostly account for the predictability effect, Baker and Bradlow [5] provided evidence for other co-existing factors, such as speech styles and information structure that could also modify acoustic cues.

Further support of the latter came from a recent study showing the combined effects of syllable-based predictability (defined as surprisal) and Lombard style on syllable duration in German [13], and from an analysis of word-final syllable duration in a German radio corpus in [1]. In the latter, syllable duration increased from low-surprisal to high-surprisal syllables, but the surprisal-induced durational adjustment was more pronounced when the syllable occurred before a weak rather than a strong intonational phrase boundary (IP). That study operationalized the strength of the IP boundary in terms of pause duration: short pause duration implied weak IP, and long pause duration strong IP. The reported interaction of surprisal and IP boundary strength suggests a possible trading relationship.

However, that study in [1] did not examine cases at an intermediate phrase (ip) boundary. Moreover, the measured syllable duration did not control for the number of syllables in a word. Since a word-final syllable in polysyllabic words is less likely to be stressed or accented than in monosyllabic words, this could have biased the measured syllable duration. Furthermore, other factors such as information structure (e.g., information status), which is known to affect duration [e.g., 5], were not factored in.

If pause provides additional time for a speaker to prepare and encode what to say next, as assumed in previous research [e.g., 10, 11], it is more likely for a pause to occur in less predictable contexts (on the assumption that speech planning is not completed before speaking). If a pause is present, this raises the second question as to whether its presence might attenuate the surprisal-induced modification of syllable duration to reduce the abrupt surge in information during signal transmission, resulting in a possible trading relationship between pause usage or pause duration and syllable duration. The present study aimed at testing the effect of contextual predictability on the absence vs. presence of pause, and the trading relationship between the pause duration and the contextual predictability-induced syllable duration, while taking into consideration the number of syllables, prosodic boundary types, information status and accenting of the host word.

To investigate any trading relationships between pause (in terms of usage and duration) and word-final syllable duration in different surprisal contexts, at different prosodic boundary types and with different information status, three hypotheses were posited: (1) more incidence of pauses in a less than in a more predictable context; (2) shorter word-final syllable duration in the presence of a pause than in its absence because the incidence of a pause and syllable duration share the same burden of minimizing any abrupt surge in information arising from a high-surprisal syllable during signal transmission; (3) pause duration negatively varies with surprisal-induced syllable duration when a pause is present.

In addition, we also expected high surprisal contexts, an intonational phrase boundary and new information status to exhibit longer syllable duration than low surprisal contexts, an intermediate phrase boundary and given information status, respectively.

2. METHOD

Analysis was based on data from the DIRNDL corpus (Discourse Information Radio News Database for Linguistic analysis), a collection of 5-hour news recordings in German from 9 speakers (5M, 4F). The corpus contained orthographic transcription, labelled information structure (e.g., lexical information status) and syntactic constituents [9]. In addition, prosodic information covering pitch accent types and prosodic boundary were annotated in the GToBI(S) framework [19].

2.1. Data

The analysed data consisted of word-final syllables in monosyllabic and polysyllabic words immediately preceding an intermediate phrase (-) or intonational phrase (%) boundary in the DIRNDL corpus, resulting in a total of 2490 items.

2.2. Language Modelling

To estimate language-based surprisal values for word-final syllables in the DIRNDL corpus, language models were first constructed from the deWaC (**de**utsche **W**eb **a**s **C**orpus) corpus [7]. The corpus consisted of web-crawled data totalling about 1.7 billion word tokens and 8 million word types from different genres, for example, newspaper articles and chat messages. During the pre-processing stage German Festival was used to remove unnecessary or duplicate document information from the raw data [20]. The normalized data were then divided into a training set (80%) and a test set (20%). The best-performing language model was based on syllable-level trigrams, after training and evaluating different language models using the SRILM toolkit [23] with Witten-Bell smoothing [26]. Surprisal values were derived from the best-performing language model and defined as the conditional probability of a unit, given two preceding units including syllable and word junctures in (1) where S = surprisal and P = probability [12].

(1) $$S(unit) = -log_2 P(unit|unit_{-1}, unit_{-2})$$

2.3. Descriptive statistics of the data

Information status		Pause (no)		Pause (yes)	
	Boundary	n	Mean (SD)	n	Mean (SD)
given	-	302	229 (75)	35	284 (97)
	%	9	277 (101)	140	249 (83)
new	-	989	243 (88)	74	296 (99)
	%	35	290 (102)	906	279 (106)

Table 1: Mean word-final syllable duration grouped by information status (given vs. new),

prosodic boundary types (intermediate [-] vs. intonational [%]) and pause occurrence (no vs. yes).

Table 1 summarizes the mean final syllable duration preceding an intermediate (-) or intonational (%) phrase boundary with lexically given or new information status when a pause is absent or present.

2.4. Statistical analysis

Syllable duration was extracted from the data using a custom Python script for analysis in R [22], using the lme4 package [6]. Surprisal values were log-transformed to adjust for skewness and then mean-centred. The statistical model included accenting (no vs. yes), information status (given vs. new), boundary type (intermediate vs. intonational), pause occurrence (absence vs. presence) and surprisal as predictors. Random structure of the model included by-speaker and by-item intercepts and by-speaker slope for predictors (subject to model convergence without overfitting). Significance testing for effects was evaluated using *F* values of type III anova which implemented Satterthwaite approximation for degrees of freedom [16]. The structure of the model was syllable duration ~ pause occurrence + boundary + information status + accenting + surprisal + information status * boundary + information status * boundary * accenting + information status * boundary * pause occurrence + boundary * pause occurrence * accenting + (pause occurrence + information status | speaker) + (1 | item).

3. RESULTS

To address hypothesis (1) regarding whether contextual predictability affects the incidence of pause, we first divided surprisal values into 2 groups with approximately equal number of observations in each (n=1246 with low surprisal and n=1244 with high surprisal); and conducted a two-proportions Z-test to compare the observed proportions of pause usage in low- vs. high-surprisal groups. The result revealed equal proportion of pause usage in low- vs. high-surprisal groups, with 562 pauses in the former and 593 in the latter (chi-squared = 1.5, df=1, p=.21).

To address hypothesis (2) as to whether a trading relationship exists between the incidence of pause and syllable duration, the lmer model results yielded significant effects of pause occurrence (F = 5.5, df = 1, p = .02*), information status (F = 8.92, df = 1, p = .004**), surprisal (F = 67.35, df = 1, p < .0001***), and accenting (F = 54.25, df = 1, p < .0001***), with two 2-way interactions: information status * boundary (F = 4.59, df = 1, p = .03*) and pause occurrence * boundary (F = 6.28, df = 1, p = .01*). No other effects or interactions reached statistical significance.

Counter to our prediction, the results did not exhibit a trading relationship between the incidence of a pause and the acoustic implementation of word-final syllable duration (Figure 1). In addition, the effect of pause occurrence on syllable duration was conditional on the types of the contiguous prosodic boundary. Word-final syllable duration was longer when a pause occurred at the intermediate phrase boundary (-) than when no pause was present (t = -4.2, df = 51, p =.0001***). Such a pause-related durational effect was not observed at the intonational phrase boundary (t= -.18, df = 258, p =.86).

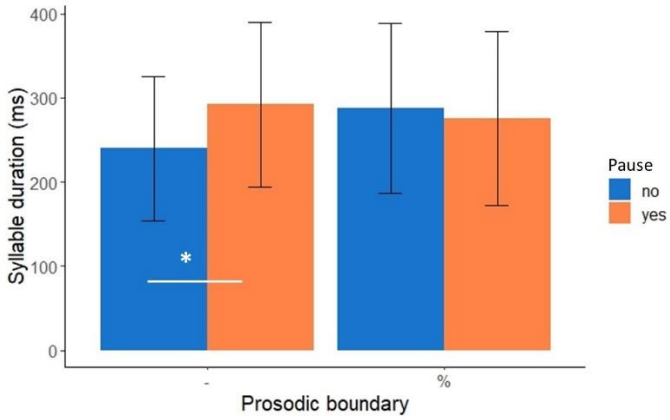

Figure 1: Mean syllable duration grouped by pause occurrence and prosodic boundaries with +/-1 SD.

Consistent with previous reports in [1], we observed a significant effect of surprisal. Information status interacted with prosodic boundary type to influence the acoustic realization of word-final syllable duration (Figure 2), with longer duration for items with lexically new than given information at an intonational phrase boundary (t=-3, df=367.3, p =.003**). Such a durational adjustment was not observed at an intermediate phrase boundary (t=-1.1, df=77.8, p =.27).

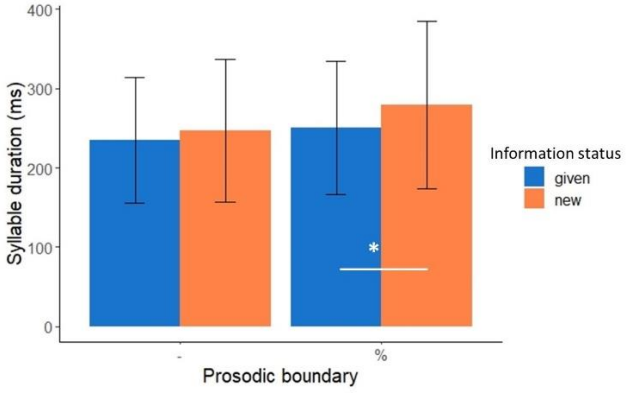

Figure 2: Mean syllable duration grouped by information status and prosodic boundaries with +/-1 SD.

To test for the existence of a trading relationship between the duration of a pause and that of word-final syllable when pause is present as stated in hypothesis (3), a series of correlations (Pearson) were carried out. Since accenting was a significant predictor in the lmer model results, correlations were conducted separately for accented and unaccented items. Figures 3 and 4 illustrate the correlation patterns for accented and unaccented items respectively. Given the relatively few tokens at an intermediate phrase boundary (-), correlations were carried out at an intonational phrase boundary. The results indicated a significant negative correlation between syllable and pause duration for accented items with new information ($t = -2.1$, $df = 209$, $p = .04$*). No other significant correlations were found.

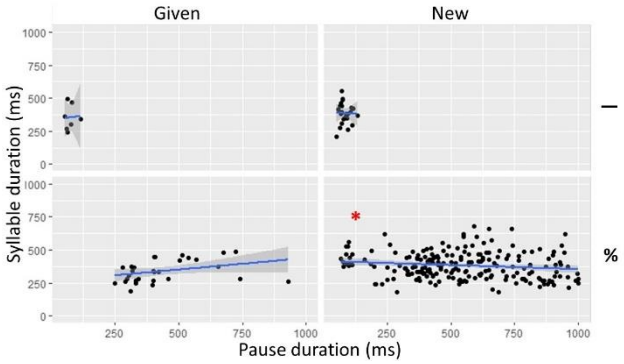

Figure 3: Scatterplots correlating pause duration (x-axis) and syllable duration (y-axis) among accented items, grouped by information status and prosodic boundaries.

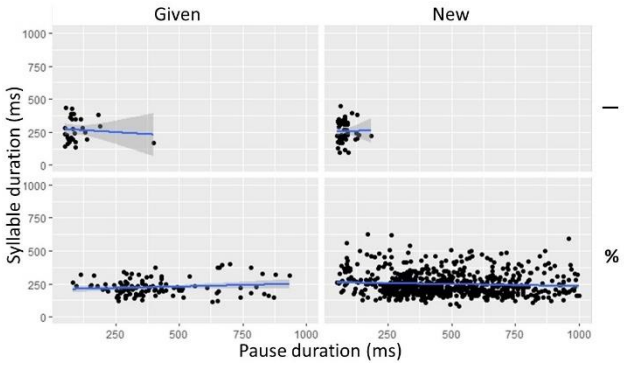

Figure 4: Scatterplots correlating pause duration (x-axis) and syllable duration (y-axis) among unaccented items, grouped by information status and prosodic boundaries.

4. DISCUSSION

Our results did not support hypothesis (1) revealing no trading relationship between the incidence of a pause and contextual predictability. This suggests that the incidence of a pause might not be related to contextual predictability in the current study of radio news which is not spontaneous speech.

Similarly, our results did not support hypothesis (2) either. There was no evidence for a trading relationship between the incidence of pause and word-final syllable durations. Word-final syllable durations were not shorter when a pause is present than when it is absent. Instead, syllable durations were long when a pause is present. Yet this pattern was found at an intermediate phrase boundary, not at an intonational phrase boundary. Since an intermediate phrase is part of an intonational phrase, more phonological planning might be needed before an intonational phrase is attained. This suggests that an intermediate phrase boundary might index a less certain location than an intonational phrase boundary. It is at such a less certain location that syllable duration was longer and the incidence of a pause more likely. This pattern (i.e., syllable duration and the incidence of a pause) might reflect the incremental nature of speech planning. That is, the phonological and phonetic specifications of a prosodic unit might not have been completed before speaking begins.

However, our results provided some evidence for hypothesis (3) that a trading relationship exists between the durations of a pause and a syllable when a pause is present for items at an intonational phrase boundary. When the duration of a word-final syllable is short, an accompanying pause duration is long. Such a trading relationship was observed for accented items, not unaccented items. The asymmetry could have arisen because the syllable duration without accenting is generally shorter than that with accenting. This durational difference might have constrained the manifestation of the trading relationship between syllable and pause duration.

5. CONCLUSION

In short, these results indicate that contextual predictability in terms of surprisal affects syllable and pause duration in a compensatory manner (i.e., in the form of a trading relationship) when a pause is present, although it does not affect pause usage (i.e., the incidence of pause). The existence of a trading relationship between pause and syllable durations suggests that speakers use these two temporal cues efficiently to alleviate any sudden surge in information density during transmission, without being redundant.

6. ACKNOWLEDGEMENTS

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

7. REFERENCES

[1] Andreeva, B., Möbius, B., & Whang, J. 2020. Effects of surprisal and boundary strength on phrase-final lengthening. *Proceedings in 10th International Conference on Speech Prosody*.

[2] Arnon, I., & Cohen Priva, U. 2013. More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, 56, 349-371.

[3] Aylett, M., & Turk, A. 2004. The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence duration in spontaneous speech. *Language and Speech*, 47, 31-56.

[4] Aylett, M., & Turk, A. 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.

[5] Baker, R. E., & Bradlow, A. R. 2009. Variability in word duration as a function of probability, speech style and prosody. *Language and Speech*, 52(4), 391-413.

[6] Bates, D., Mächler, M., Bolker B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, 1-48.

[7] Baroni, M., Bernardini, S., Ferraresi, A., Zanchetta, E. 2009. The WaCky Wide Web: A collection of very large linguistically processed web-crawled corpora, *Language Resources and Evaluation*, 43, 209-226.

[8] Brandt, E., Möbius, B., & Andreeva, B. 2021. Dynamic formant trajectories in German read speech: Impact of predictability and prominence. *Frontiers in Communication*, 6, doi: 10.3389/fcomm.2021.643528.

[9] Eckart, K., Riester, A., Schweitzer, K. 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.

[10] Goldman-Eisler, F. 1958a. Speech production and the predictability of words in context. *Quarterly Journal of Experimental Psychology* 10, 96-106.

[11] Goldman-Eisler, F. 1958b. The predictability of words in context and the length of pauses in speech. *Language and Speech*, 1, 226-231.

[12] Hale, J. 2016. Information-theoretical complexity metrics. *Language and Linguistic Compass*, 10, 397-412.

[13] Ibrahim, O., Yuen, I., van Os, M., Andreeva, B., & Möbius, B. 2022. The combined effects of contextual predictability and noise on the acoustic realization of German syllables. *Journal of the Acoustical Society of America*, 162(2), 911-920. Doi:10.1121/10.0013413

[14] Krivokapić, J. Styler, W., Parrell, B. 2020. Pause postures: the relationship between articulation and cognitive processes during pauses. *Journal of Phonetics*, 79, doi: 100953.

[15] Krivokapić, J., Styler, W., Byrd, D. 2022. The role of speech planning in the articulation of pauses. *Journal of the Acoustical Society of America*, 151, 402-413.

[16] Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. 2017. lmerTest: Tests in linear mixed effects models: R package version 2.0-33

[17] Maclay, H., Osgood, C. E. 1959. Hesitation phenomena in spontaneous English speech. *Word*, 15, 19-44.

[18] Malisz, Z., Brandt, E., Möbius, B., Yoon, M. O., Andreeva, B. 2018. Dimensions of segmental variability: interaction of prosody and surprisal in six languages. *Frontiers in Communication/Language Sciences*, 3, 1-18.

[19] Mayer, J. 1995. Transcription of German Intonation – the Stuggart system. Institue of Natural Language Processing. *University of Stuggart Tech. Rep*. doi: doi:https://www.ims.uni-stuttgart.de/documents/arbeitsgruppen/ehemalig/ep-dogil/STGTsystem.pdf

[20] Möhler, G., Schweitzer, A., Breitenbücher, M., Barbisch, M. 2000. IMS German Festival (version 1.2-os). University of Stuggart: Institute für Maschinelle Sprachverarbeitung (IMS) retrieved 2020-01-02.

[21] Piantadosi, S. T., Tily, H. J. Gibson, E. 2011. Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108, 3526-3529.

[22] R Core Team. 2022. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria).

[23] Stolcke, A. 2002. SRILM – an extensible language modeling toolkit. *Proceedings of Interspeech, Denver, Colorado.* vol. 2, 901-904.

[24] Swerts, M. 1997. Prosodic features at discourse boundaries of different strength. Journal of the Acoustical Society of America, 101, 514-521.

[25] Tisljár-Szabó, E., Pléh, C. 2014. Ascribing emotions depending on pause length in native and foreign language speech. *Speech Communication*, 56, 35-48.

[26] Witten, I. H., Bell, T. C. 1991. The zero-frequency problem: Estimating the probabilities of novel events in adaptive text compression. *IEEE Transactions on Information Theory*, 37, 1085-1094.

[27] Yang, L. 2004. Duration and pauses as cues to discourse boundaries in speech. In Proceedings of Speech Prosody, pp. 35-48